An Exploration in the Theory of Optimum Income Taxation

J. A. Mirrlees

The Review of Economic Studies, Vol. 38, No. 2. (Apr., 1971), pp. 175-208.

Stable URL:

http://links.jstor.org/sici?sici=0034-6527%28197104%2938%3A2%3C175%3AAEITTO%3E2.0.CO%3B2-V

The Review of Economic Studies is currently published by The Review of Economic Studies Ltd..

An Exploration in the Theory of Optimum Income Taxation[1,2]

J. A. MIRRLEES

Nuffield College, Oxford

1. INTRODUCTION

One would suppose that in any economic system where equality is valued, progressive income taxation would be an important instrument of policy. Even in a highly socialist economy, where all who work are employed by the State, the shadow price of highly skilled labour should surely be considerably greater than the disposable income actually available to the labourer. In Western Europe and America, tax rates on both high and low incomes are widely and lengthily discussed[3]: but there is virtually no relevant economic theory to appeal to, despite the importance of the tax.

Redistributive progressive taxation is usually related to a man's income (or, rather, his estimated income). One might obtain information about a man's income-earning potential from his apparent I.Q., the number of his degrees, his address, age or colour: but the natural, and one would suppose the most reliable, indicator of his income-earning potential is his income. As a result of using men's economic performance as evidence of their economic potentialities, complete equality of social marginal utilities of income ceases to be desirable, for the tax system that would bring about that result would completely discourage unpleasant work. The questions therefore arise what principles should govern an optimum income tax; what such a tax schedule would look like; and what degree of inequality would remain once it was established.

The problem seems to be a rather difficult one even in the simplest cases. In this paper, I make the following simplifying assumptions:

(1) Intertemporal problems are ignored. It is usual to levy income tax upon each year's income, with only limited possibilities of transferring one year's income to another for tax purposes. In an optimum system, one would no doubt wish to relate tax payments to the whole life pattern of income,[4] and to initial wealth; and in scheduling payments one would wish to pay attention to imperfect personal capital markets and imperfect foresight. The economy discussed below is timeless. Thus the effects of taxation on saving are ignored. One might perhaps regard the theory presented as a theory of " earned income " taxation (i.e. non-property income).

(2) Differences in tastes, in family size and composition, and in voluntary transfers, are ignored. These raise rather different kinds of problems, and it is natural to assume them away.

[3] Discussions on (usually) orthodox lines, including many important points neglected in the present paper, can be found in [7], [1], [5, Chapters 5, 7, 8], and [6, Chapters 11 and 12]. [2] is close in spirit to what is attempted here.
[4] Cf. [7, Chapter 6].

(3) Individuals are supposed to determine the quantity and kind of labour they provide by rational calculation, corresponding to the maximization of a utility function, and social welfare is supposed to be a function of individual utility levels. It is also supposed that the quantity of labour a man offers may be varied within wide limits without affecting the price paid for it. The first assumption may well be seriously unrealistic, especially at higher income levels, where it does sometimes appear that there is consumption satiation and that work is done for reasons barely connected with the income it provides to the " labourer ".

(4) Migration is supposed to be impossible. Since the threat of migration is a major influence on the degree of progression in actual tax systems, at any rate outside the United States, this is another assumption one would rather not make.[1]

(5) The State is supposed to have perfect information about the individuals in the economy, their utilities and, consequently, their actions. In practice, this is certainly not the case for certain kinds of income from self-employment, in particular work done for the worker himself and his family; and in some countries, the extent of uncertainty about incomes is very great. Yet it seems doubtful whether the neglect of this uncertainty is a simplification of much significance.

(6) Various formal simplifications are made to render the mathematics more manageable: there is supposed to be one kind of labour (in a special sense to be explained below); there is one consumer good; welfare is separable in terms of the different individuals of the economy, and symmetric—i.e. it can be expressed as the sum of the utilities of individuals when the individual utility function (the same for all) is suitably chosen).

(7) The costs of administering the optimum tax schedule are assumed to be negligible.

In sections 2-5, the more general properties of the optimum income-tax schedule, and the rules governing it, are discussed. The treatment is not rigorous. Nevertheless a reader who wants to avoid mathematical details can omit the last page or two of section 3, and will probably want to glance through section 4 rather rapidly. In section 6, I begin the discussion of special cases. The mathematical arguments in sections 6-8 are frequently complicated. If the reader goes straight to section 9, where numerical results are presented and discussed, he should not find the omission of the previous sections any handicap. He may, nevertheless, find it interesting to look at the results and conjectures presented at the beginning of section 7, and at the diagrams for the two cases discussed in section 8.

Rigorous proofs of the main theorems will be given in a subsequent paper, [4].

2. MODEL AND PROBLEM

Individuals have identical preferences. We shall suppose that consumption and working time enter the individual's utility function. When consumption is x and the time worked y, utility is

$$u(x, y).$$

x and y both have to be non-negative, and there is an upper limit to y, which is taken to be 1. In fact, it is assumed that: u is a strictly concave, continuously differentiable, function (strictly) increasing in x, (strictly) decreasing in y, defined for $x > 0$ and $0 \leqq y < 1$. u tends to $-\infty$ as x tends to 0 from above or y tends to 1 from below.

The usefulness of a man's time, from the point of view of production, is assumed to vary from person to person. To each individual corresponds a number n such that the quantity of labour provided, per unit of his time, is n. If he works for time y, he provides a quantity of labour ny. There is a known distribution of skills, measured by the parameter n, in the population. The number of persons with labour parameter n or less is $F(n)$. It

---

[1] The relation of optimum tax schedules to propensities to migrate is discussed in another paper under preparation.

will be assumed that F is differentiable, so that there is a density function for ability, $f(n) = F'(n)$. Call an individual whose ability-parameter is n an n-man.

The consumption choice of an n-man is denoted by (x_n, y_n). Write $z_n = ny_n$ for the labour he provides. Then the total labour available for use in production in the economy is

$$Z = \int_0^\infty z_n f(n) dn, \qquad \ldots(1)$$

and the aggregate demand for consumer goods is

$$X = \int_0^\infty x_n f(n) dn. \qquad \ldots(2)$$

In order to avoid the possibility of infinite labour supply, I assume that

$$\int_0^\infty n f(n) dn < \infty. \qquad \ldots(3)$$

Each individual makes his choice of (x_n, y_n) in the light of his budget constraint. Using an income tax, the government can arrange that a man who supplies a quantity of labour z can consume no more than $c(z)$ after tax: the government can choose the function c arbitrarily. It makes sense to impose the restriction on the government's choice of c, that c be upper semi-continuous, for then all individuals have available to them consumption choices that maximize their utility, subject to the budget constraint[1]:

$$(x_n, y_n) \text{ maximizes } u(x, y) \text{ subject to } x \leqq c(ny). \qquad \ldots(4)$$

Notice that (x_n, y_n) may not be uniquely determined for every n.[2] I write:

$$u_n = u(x_n, y_n). \qquad \ldots(5)$$

Proposition 1. *There exists a number $n_0 \geqq 0$ such that*

$$y_n = 0 \quad (n \leqq n_0),$$
$$y_n > 0 \quad (n > n_0). \qquad \ldots(6)$$

Proof. If $m < n$, and $y_m > 0$, $u[c(my_m), y_m] < u\left[c\left(n \cdot \dfrac{m}{n} y_m\right), \dfrac{m}{n} y_m\right] \leqq u_n$. Consequently, $y_m = 0$ if $y_n = 0$, since then $y_m = 0$ gives the utility u_n to n-man. Thus

$$n_0 = \inf [n \mid y_n > 0]$$

has the desired properties.‖

Proposition 2. *Any function[3] of n, (x_n, y_n), that satisfies (4) for some upper semi-continuous function c also satisfies (4) for some non-decreasing, right-continuous function c'.*

---

[1] To say that c is upper semi-continuous means that
$$\limsup c(z_i) = c(z) \text{ when } \lim_{i \to \infty} z_i = z.$$
If
$$u_n = \sup \{u(x, y) \mid x \leqq c(ny)\}, \text{ and } u(x_i, y_i) \to u_n, x_i \leqq c(ny_i)$$
we can suppose that $x_i \to x$ and $y_i \to y$ (since $\{y_i\}$ and therefore $\{x_i\}$ is bounded). By the upper semi-continuity of c,
$$x \leqq \limsup c(ny_i) = c(ny);$$
and by the continuity of u, $u(x, y) = \lim u(x_i, y_i) = u_n$. Therefore the supremum is attained.

[2] In other words, we have a *correspondence*, providing a set of utility maximizing choices for n-men. It arises when the consumption function c coincides with the indifference curve for part of its length. It is convenient nevertheless to use the notation of the text, despite its suggestion that we are dealing with a function.

[3] It is easy to see that the result is true for a correspondence also.

Proof.   Define $c'(z) = \sup\limits_{z' \le z} c(z')$.   If $x'_n \le c'(ny'_n)$, then, for any $\varepsilon > 0$, there exists $y''_n \le y'_n$ such that $x'_n - \varepsilon \le c(ny''_n)$. Thus $u(x'_n - \varepsilon, y''_n) \le u_n$, which implies, since u is a decreasing function in y, that $u(x'_n - \varepsilon, y'_n) \le u_n$. Letting $\varepsilon \to 0$, $u(x'_n, y'_n) \le u_n$. It follows that (x_n, y_n) maximizes u subject to $x \le c'(ny)$.

c' is clearly a non-decreasing function of z. To prove that it is right-continuous, take a decreasing sequence $z^i \to z$. $c'(z^i)$ is a non-increasing sequence, and therefore tends to a limit, which is not less than $c'(z)$. If it is equal to $c'(z)$, there is no more to prove. Suppose it is greater. Then for some $\varepsilon > 0$ each $c'(z^i) > c'(z) + \varepsilon$. Therefore, there exists a sequence $(\bar{z}^i)$ such that $\bar{z}^i \le z^i$ and $c'(z^i) \ge c(\bar{z}^i) > c'(z) + \varepsilon$. The second inequality implies that $\bar{z}^i > z$. Thus $\bar{z}^i \to z$. Yet $\limsup c(\bar{z}^i) > c(z)$, which contradicts upper semi-continuity. Thus in fact, c is right-continuous. ‖

This proposition says that the marginal tax rate may as well be not greater than 100 per cent. We shall consider later whether it should be positive.

The government chooses the function c so as to maximize a welfare function

$$W = \int_0^\infty G(u_n) f(n) dn. \qquad \qquad ...(7)$$

I use the function G here, rather than writing u_n alone, because I shall later want to devote special attention to the case $u_{xy} = 0$ (when u can be written as the sum of a function depending only on x and a function depending only on y). In maximizing welfare, the government is constrained by production possibilities: it must be possible to produce the consumption demands, X, arising from its choice of c, with labour input no greater than Z. The production constraint is written

$$X \le H(Z). \qquad \qquad ...(8)$$

We have not yet fully specified the possibilities available to the government, since, if (x_n, y_n) is not uniquely defined, it is not clear whether the government or the consumer is allowed to choose the particular utility-maximizing point. Perhaps it is reasonable to suppose that the government can choose, and that the necessity for market-clearing will make its choices actual. But it will turn out that the issue is of no significance when we make the following assumption, as we shall:

(A) y_n is uniquely defined for all n except for a set of measure 0.

Thus the class of functions c from which the government chooses is further restricted by the requirement that the function lead to choices satisfying (A). It will appear in due course that (A) is satisfied for *all* functions c in the particular cases we shall be most concerned with.

3. NECESSARY CONDITIONS FOR THE OPTIMUM

On the assumption that an optimum for our problem exists, we shall now obtain conditions that it must satisfy. The mathematical argument will not be rigorous. To do the analysis properly, one must attend to a number of rather tricky points. Since these technical details tend to obscure the main lines of the argument, rigorous proofs will be presented separately, in the continuation of this paper. The nature of these neglected difficulties will be discussed briefly in the next section.

The key to a reasonably neat solution of the problem is to find a convenient expression of the condition that each man maximizes his utility subject to the imposed " consumption function " c. If we suppose that c is differentiable, the derivative of $u[c(ny), y]$ with respect to y must be zero. Denoting the derivative of u with respect to its first and second arguments by u_1 and u_2, respectively, we have

$$u_1 nc'(ny) + u_2 = 0. \qquad \qquad ...(9)$$

Recollect that u_n is the utility of n-man. Then a straightforward calculation, using the first-order condition (9), yields

$$\frac{du_n}{dn} = u_1 y c' = -\frac{y u_2}{n}. \qquad \qquad ...(10)$$

(The expressions on the right are, of course, alternative expressions for the *partial* derivative of u with respect to n, evaluated at the maximum. The case where n enters u in a more general manner can be analyzed by using this more general equation. We shall return to this point later.)

Our problem is to maximize w subject to the constraint of the production function, $X \leqq H(Z)$, the differential equation (10), and the definition $u_n = u(x_n, y_n)$. Those who are familiar with the Pontriyagin Maximum Principle will see that this is a form of problem fairly suitable for treatment by it. Shadow prices p and w have to be introduced for X and Z. Then we would like to maximize

$$W - pX + wZ = \int [G(u_n) - px_n + wy_n n] f(n) dn \qquad \qquad ...(11)$$

subject to (10). u_n is to be regarded as the state variable, y_n (say) as the control variable, while x_n is determined as a function of u_n and y_n from the equation $u_n = u(x_n, y_n)$. The Hamiltonian is

$$M = G[(u_n) - px_n + wy_n n] f(n) - \phi_n \frac{y_n u_2}{n},$$

where ϕ_n is a function of n satisfying the differential equation

$$\frac{d\phi}{dn} = -\frac{\partial M}{\partial u}$$

$$= -\left[G'(u_n) - \frac{p}{u_1} \right] f(n) + \phi \frac{y_n u_{12}}{n u_1}. \qquad \qquad ...(12)$$

y_n should then be chosen so as to maximize M:

$$\left[wn + \frac{p u_2}{u_1} \right] f(n) + \phi_n \frac{\psi_y}{n} = 0, \qquad \qquad ...(13)$$

where the function $\psi(u, y)$ is defined by

$$\psi(u, y) = -y u_2(x, y), \quad u = u(x, y), \qquad \qquad ...(14)$$

and ψ_y is its partial derivative with respect to y. (Notice, at the same time, that

$$\psi_u = -y u_{12}/u_1.)$$

Equation (12) can now be integrated to obtain an expression for ϕ_n; which, when substituted in (13), provides us with an equation to be satisfied by the optimum we seek. Before going on to use this equation, however, we shall derive it in a different way, by a more explicit use of the methods of the calculus of variations. The use of the Maximum Principle has a number of serious disadvantages. It does not show us how to obtain certain important supplementary conditions on the optimum. The analysis provides no hint as to how it could be made rigorous. It does not provide any insight into the kind of maximization that is going on. When we have done a more explicit variational analysis, we shall be better able to see where the logical holes are, and to understand why things come out the way they do.

For this purpose, I prefer to write (10) in integrated form:

$$u_n = - \int_0^n y_m u_2(x_m, y_m) \frac{dm}{m} + u(c(0), 0),$$

$$= \int_0^n \psi(u_m, y_m) \frac{dm}{m} + u_0, \qquad \qquad ...(15)$$

using the notation ψ introduced above, and denoting the utility allowed to a man who does no work by u_0. Suppose first that ψ is independent of u (corresponding to the sepcial case $u_{12} = 0$). If we consider a variation from the optimum which changes the functions u_n and y_n by " small " variations δu_n and δy_n, we deduce from (15) that these variations must be related by

$$\delta u_n = \int_0^n \psi_y \delta y_m \frac{dm}{m} + \delta u_0. \qquad \qquad ...(16)$$

This variation will bring about changes in W, X, and Z. As before, introduce shadow prices (in terms of welfare) for X and Z. Then the variation must leave (11) stationary:

$$0 = \delta \int [G(u_n) - px_n + wy_n n] f(n) dn$$

$$= \int \int \left[G'(u_n) \delta u_n - p \left(\frac{1}{u_1} \delta u_n - \frac{u_2}{u_1} \delta y_n \right) + w \delta y_n n \right] f(n) dn, \qquad ...(17)$$

where the variation in x is calculated as follows:

$$\delta u_n = \delta u(x_n, y_n) = u_1 \delta x_n + u_2 \delta y_n. \qquad \qquad ...(18)$$

It remains to substitute (16) in (17), yielding,

$$0 = \int_0^\infty \left\{ \left[G'(u_n) - \frac{p}{u_1} \right] \left[\int_0^n \psi_y \delta y_m \frac{dm}{m} + \delta u_0 \right] + \left[wn + p \frac{u_2}{u_1} \right] \delta y_n \right\} f(n) dn$$

$$= \int_0^\infty \left\{ \int_n^\infty \left[G'(u_m) - \frac{p}{u_1} \right] f(m) dm \cdot \frac{\psi_y}{n} + \left(wn + p \frac{u_2}{u_1} \right) f(n) \right\} \delta y_n dn$$

$$+ \int_0^\infty \left[G'(u_n) - \frac{p}{u_1} \right] f(n) dn \cdot \delta u_0. \qquad ...(19)$$

The second equation is obtained by inverting the order of integration in the double integral.[1] (19) is to be satisfied for all possible variations of the function y_n, and the number u_0. Since u_0 can be either increased or decreased at the optimum (if, as is to be expected in general, some people will do no work at the optimum),

$$\int_0^\infty \left[G'(u_n) - \frac{p}{u_1} \right] f(n) dn = 0 \qquad \qquad ...(20)$$

at the optimum.

[1] The double integral is

$$\int_0^\infty \left[G'(u_n) - \frac{p}{u_1} \right] f(n) \int_0^n \psi_y \delta y_m \frac{dm}{m} \cdot dn.$$

The region over which the integration takes place is defined by $0 \leq m \leq n$. Thus, when the order of integration is inverted, n ranges between m and ∞ for given m. The integral can therefore be written

$$\int_0^\infty \int_m^\infty \left[G' - \frac{p}{u_1} \right] f(n) dn \cdot \psi_y \delta y_m \cdot \frac{dm}{m},$$

which is seen to justify (19) on permuting the symbols m and n.

If all variations in y_n were possible—and this is a question we shall take up shortly—we could also claim that the expression within curly brackets ought to be zero:

$$\left(wn + p\frac{u_2}{u_1}\right)f(n) = \frac{\psi_y}{n}\int_n^\infty\left[\frac{p}{u_1} - G'(u_m)\right]f(m)dm. \qquad ...(21)$$

It should be noticed that this equation will only be valid for $n \geqq n_0$: it does not apply to n for which $y_n = 0$ (except n_0) because, there, not all variations of the function y_n are possible, since y_n cannot be negative.

Finally we know that the marginal product of labour should be equal to the shadow wage:

$$pH'(Z) = w. \qquad ...(22)$$

These equations, (20) and (21), have been worked out under the special assumption that ψ is independent of u. In the more general case, we have to replace (16) by

$$\delta u_n = \int_0^n T_{mn}\psi_y\delta y_m\frac{dm}{m} + \delta u_0, \qquad ...(23)$$

where

$$T_{mn} = \exp\int_m^n\psi_u\frac{dm'}{m'}. \qquad ...(24)$$

To show this, we can go back to the differential equation (10). Applying the variation, we obtain from it,

$$\frac{d}{dn}\delta u_n = \frac{1}{n}\psi_u\delta u_n + \frac{1}{n}\psi_y\delta y_n. \qquad ...(25)$$

This is a first order linear equation, and can therefore be solved by the standard method to give the solution (23).

Having replaced (16) by (23), we can now go through the rest of the calculation as before. We find that (20) is generalized into

$$\int_0^\infty\left[G'(u_n) - p/u_1\right]T_{0n}f(n)dn = 0; \qquad ...(26)$$

while (21) becomes

$$(wn + pu_2/u_1)f(n) = \frac{\psi_y}{n}\int_n^\infty\left[p/u_1 - G'(u_m)\right]T_{nm}f(m)dm. \qquad ...(27)$$

Notice that we have T_{nm} here, although it was T_{mn} that appeared in (23).

If these equations are correct, the two integral equations, (15) and (27) may be thought of as determining the two functions u_n and y_n, given the three parameters $u_0, w,$ and p. The values of these parameters are fixed by the three equations (26), (22), and (8). We have enough relations to determine the optimum tax schedule, since the function c can be determined once we know u_n and y_n.

4. NECESSARY CONDITIONS: A COMPLETE STATEMENT

The argument used to derive these conditions for the optimum tax schedule had a number of weak points. It is indeed unlikely that the relationships derived above hold in general. Among the weak points of the argument, notice that

(i) the existence of the shadow prices p and w was assumed without proof;
(ii) the optimum tax schedule, and the resulting functions $x_n, y_n,$ and u_n were assumed to be differentiable;
(iii) the application of the variation was quite heuristic; and
(iv) no justification was provided for assuming that the function y_n could be varied arbitrarily (for $n > n_0$).

M

I shall not comment on (i) and (iii), which, though important, are technical matters: they can be justified. (ii) is not satisfied in general: there was no reason to suppose that it would be. When (ii) is not satisfied, the first-order condition, (9), for maximization of utility ceases to be meaningful. Finally, (iv) is never justified. The function y_n is derived from the imposition of the consumption function c, and we have no *a priori* information about it. We must expect that some conceivable functions y_n can never arise from the imposition of a consumption function. The class of possible y-functions is no doubt quite complicated in certain cases. Fortunately it is possible to specify that class quite simply in the realistic cases, and it is then possible to use the variational argument rigorously.

Problem (ii) is dealt with in the rigorous analysis by depending on equation (15) instead of the differential first-order condition (9). It is a remarkable fact that this condition holds if and only if the various functions arise from utility-maximization under an imposed consumption function, even when that function is not differentiable. For proof, the reader is referred to [4].

To deal with problem (iv), we have to restrict the class of utility functions considered. We assume that

(B) $V(x, y) = -yu_2/u_1$ is an increasing function of y for each $x>0$ (and bounded in $0 \leq x \leq \bar{x}, 0 \leq y \leq \bar{y}$ for any $\bar{x}<\infty$ and $\bar{y}<1$).

It will be noticed that this is an assumption about preferences, not just about the form of the utility function used to represent preferences. The second part of the assumption is readily acceptable. The first, and main part of the assumption holds if and only if, for a given level of consumption x, a one per cent increase in the amount of work done requires a larger increase in consumption to maintain the same utility level, the greater is the amount of work being done. It is equivalent to assuming that (in the absence of taxation) the consumer's demand for goods is an increasing function of the real wage rate (at any given non-wage income.[1] Few individuals appear to have preferences violating (B), and intuitively it is rather plausible. We shall later use the fact that (B) holds if preferences can be represented by an additive utility function. (It will be noticed that, as $y \to 1$, $V \to +\infty$, so that the assumption must hold for some ranges of y.) If the assumption does not hold, the theory of optimum taxation is more complicated.

The point of the assumption is indicated in

Theorem 1. *Under Assumption (B), $z_n = ny_n$ maximizes utility for every n under some consumption function c if and only if*

(i) z_n *is a non-decreasing function defined for $n>0$;*

(ii) $0 \leq z_n < n$ *for all $n>0$.*

---

[1] This equivalence is fairly obvious from an indifference curve diagram. For a formal proof that (B) implies that consumption is an increasing function of the wage rate, let w be the wage rate, and m non-labour income (both measured in terms of goods). (B) states that wy, regarded as a function of x and y, is an increasing function of y. Write x and y as functions of w and m, putting $x = x(w, m)$, $y = y(w, m)$ and $x' = x(w', m)$, $y' = y(w', m)$ where $w' > w$. I shall show that $x' > x$. To do this, choose w'' and m'' such that $x'' = x(w'', m'') = x$, and

$$y'' = y(w'', m'') = \frac{w}{w'}y.$$

Since $x'' - w'y'' = m$, (x', y') is preferred to (x'', y''); and therefore

$$x' - x > w''(y' - y'')$$
$$= \frac{w''}{w'}(w'y' - w'y'') = \frac{w''}{w'}(w'y' - wy)$$
$$= \frac{w''}{w'}(x' - x),$$

since $x' - w'y' = m = x - wy$. This implies, with our assumption $w'' < w'$, that $x' > x$. The converse proposition can be proved by reversing the steps.

For a rigorous proof of this theorem, the reader is referred to [4]. For a heuristic justification, suppose that z_n is differentiable, and that c is twice differentiable. The first order condition, (9), can be written

$$\frac{\partial}{\partial z} u(c(z), z/n) = \frac{u_1}{z} [zc'(z) - V(c(z), z/n)] = 0. \qquad \qquad \ldots(28)$$

Furthermore, we have the second-order condition, that the derivative is non-increasing at z_n. Since it is zero there, this is also true when we drop the positive factor u_1/z. In other words,

$$\frac{\partial}{\partial z} [zc'(z) - V(c(z), z/n)] \leqq 0, \text{ at } z = z_n. \qquad \qquad \ldots(29)$$

Now differentiate the equation $z_n c'(z_n) - V(c(z_n)z_n/n) = 0$ with respect to n:

$$\frac{\partial}{\partial z} [zc' - V]|_{z = z_n} \cdot \frac{dz_n}{dn} = -V_y(c(z_n), z_n/n)z/n^2. \qquad \qquad \ldots(30)$$

It follows from (29) and assumption (B) that

$$\frac{dz_n}{dn} > 0 \qquad \qquad \ldots(31)$$

unless $z_n = 0$. In fact z_n is strictly increasing when $n > n_0$ and c is differentiable; a corner in c causes z_n to be constant for a range of values of n. (An indifference curve diagram makes this clear.) Condition (ii) of the theorem clearly has to be satisfied by the utility maximizing choice.

To prove that a suitable consumption function exists for a given z-function satisfying the two conditions, one defines c by the first-order condition (28). (30) then shows (nearly) that the second-order condition for a maximum is satisfied. This does not yet prove global maximization of utility, but that also is true.

It should be noticed that, as a corollary of Theorem 1, condition (A) holds when condition (B) holds, for z_n is shown to be non-decreasing even if it is a correspondence. It therefore takes a single value for all but a countable set of values of n. *A fortiori*, condition (A) is satisfied in this case.

Theorem 1 at once implies that z_n and therefore also x_n are non-decreasing functions when the optimum tax schedule is imposed. Furthermore, it shows us quite straightforwardly what changes in the function y_n we are allowed to contemplate when applying the variational argument that allowable small changes should make only a second-order difference to the maximand. The rigorous argument is still complicated, in part because one has to allow for the possibility that z_n is constant over some intervals, and discontinuous at some values of n. The full statement of the result, which is proved in [4], is as follows:

Theorem 2. *If preferences satisfy assumption (B) and (u_n, x_n, y_n) arise from optimum income taxation, then*

(i) $z_n = ny_n$ *is a non-decreasing function of n;*

(ii) $u_n = u_0 - \int_0^n [y_m u_2(x_m, y_m)/m]dm \quad (n \geqq 0);$ $\qquad \qquad \ldots(32)$

(iii) *at all points of increase of z_n (i.e., where $z_n > z_{n'}$, for all $n' < n$, or $z_n < z_{n'}$, for all $n' > n$)*

$$A_n \equiv [w + u_2^{(n)}/nu_1^{(n)}]f(n) - \frac{\psi_y}{n^2} \int_n^\infty \left[\frac{1}{u_1^{(m)}} - \lambda G'(u_m)\right] T_{nm}f(m)dm = 0, \qquad \ldots(33)$$

where superscripts " (n) ", etc. indicate that the function is evaluated at n-man (etc.)'s utility-maximizing choice, and

$$\psi_y = -u_2^{(n)} - y_n u_{22}^{(n)} + y_n u_2^{(n)} u_{12}^{(n)}/u_1^{(n)}, \qquad \qquad ...(34)$$

$$T_{nm} = \exp\left[-\int_n^m y_{m'} u_{12}(x_{m'}, y_{m'})/u_1(x_{m'}, y_{m'}).dm'\right]; \qquad ...(35)$$

(iv) If $n \in [n_1, n_2]$, where z is constant on $[n_1, n_2]$, and $[n_1, n_2]$ is a maximal interval of constancy for z,

$$\int_{n_1}^n A_m dm \geqq 0, \quad \int_n^{n_2} A_m dm \leqq 0; \qquad \qquad ...(36)$$

(v) If z is discontinuous at n, $\bar{y}_n$ is defined to be $\lim_{m \to n-} y_m$, $\bar{x}_n$ is defined by

$$u(\bar{x}_n, \bar{y}_n) = u_n = u(x_n, y_n),$$

and $\bar{u}_1$, etc., denote u_1 evaluated at $\bar{x}_n$, $\bar{y}_n$, while u_1, etc., denote evaluation at x_n, y_n,

$$\frac{(wy_n - x_n/n) - (w\bar{y}_n - \bar{x}_n/n)}{\bar{y}_n \bar{u}_2 - y_n u_2} = \frac{w + u_2/nu_1}{\psi_y} = \frac{w + \bar{u}_2/n\bar{u}_1}{\bar{\psi}_y}. \qquad ...(37)$$

If ψ_y is a non-decreasing function of y for constant u, z_n is continuous for all n.

(vi) $$\int_0^\infty \left[\frac{1}{u_1} - \lambda G'(u_m)\right] T_{0m} f(m) dm = 0, \qquad \qquad ...(38)$$

(vii) $X = H(Z)$, $\qquad \qquad ...(39)$

$w = H'(Z)$. $\qquad \qquad ...(40)$

It will be noticed that in this statement w is the commodity shadow wage rate (w/p in the earlier notation), while λ ($1/p$ in the previous notation) is the inverse of the marginal social utility of commodities (national income). The second part of (v) should be particularly noted, since we are quite likely to be willing to assume that ψ_y is a non-decreasing function of y, and it is a great advantage not to have to worry about possible discontinuities in z_n. It does not seem possible, unfortunately, to delimit a class of cases in which one can be sure that $[0, n_0]$ will be the only interval of constancy for z. It should be mentioned that, when ψ_y is not non-decreasing, and the equations (37) may possibly apply, the conditions of Theorem 1 may define more than one candidate for optimality, and then only direct comparison of the welfare generated by the alternative paths so defined will solve the problem.

5. INTERPRETATION

If n is not in an interval of constancy for z, and $c(.)$ is therefore a differentiable function at z_n, the first-order condition (9) applies. It can be written

$$-u_2/nu_1 = c'(z). \qquad \qquad ...(41)$$

If we denote the marginal tax rate, $\dfrac{d}{d(wz)}[wz - c(z)]$, by θ, we have

$$w\theta = \frac{d}{dz}[wz - c(z)] = w + u_2/nu_1$$

$$= \frac{\psi_y}{n^2 f(n)} \int_n^\infty \frac{1 - \lambda G' u_1}{u_1} T_{nm} f(m) dm, \qquad ...(42)$$

by (33). (42) suggests the considerations that should influence the magnitude of the marginal tax rate. First, it can tell us something about the sign of θ: we already know that θ will not be greater than 1, but we were not previously able to say anything about its sign. Of course, we expect that it will not usually be negative. Using (42) and the conditions in Theorem 1, we can establish this rigorously.

Note first that $1 - \lambda G' u_1$ is a non-decreasing function of n, since x_n is a non-decreasing function of n, and $\dfrac{\partial}{\partial x} G = G' u_1$ a decreasing function of x. If $1 - \lambda G' u_1$ were always positive or always negative, Equation (38) could not be satisfied. Therefore

$$\int_n^\infty \frac{1}{u_1} (1 - \lambda G' u_1) T_{nm} f(m) dm$$

is increasing in n for n less than some $\bar{n}$, and decreasing for $n > \bar{n}$; but in any case positive for $n > \bar{n}$. (Here we use the properties $u_1 > 0$, $T_{mn} > 0$.) Since the integral is zero when $n = 0$, it is non-negative for all n. Consequently the marginal tax rate is non-negative at all points of increase of z. If n is not a point of increase of z, c is not differentiable at z_n. It is easily seen that, if $[n_1, n_2]$ is a maximal interval of constancy of z, $-u_2/nu_1$ is equal to the left derivative of c at n_1, and the right derivative at n_2. Thus both the " right " and " left " marginal tax rates are non-negative in this case. Summarizing:

Proposition 3.[1] *If assumption (B) is satisfied, $wz - c(z)$ (the " tax function ") is a non-decreasing function for all z that actually occur (and may therefore be taken to be a non-decreasing function for all z).*

Having established that the integral in Equation (42) is non-negative for all n, we can see that the marginal tax rate will be greater if there are relatively few n-men than otherwise; or if the utility-value of work, $-yu_y$, is more sensitive to work done (utility being held constant); or if n is closer to $\bar{n}$, the value of n at which $1 = \lambda G' u_1$ (and the integral is therefore a maximum). If f is a single-peaked distribution, the first consideration suggests that marginal tax rates should be greatest for the richest and the poorest; but the last consideration tells the other way.

In any case, it is important to note than n_0, the largest n for which $y_n = 0$, may be quite large: if the number who do not work in the optimum regime is large, the marginal tax rate may not be high at zero income. Explicitly, we can rewrite Equation (38) in the form

$$\left[\frac{1}{u_1(x_0, 0)} - \lambda G'(u_0) \right] F(n_0) + \int_{n_0}^\infty \left[\frac{1}{u_1} - \lambda G' \right] T_{n_0 m} f(m) dm = 0 \qquad \text{...(43)}$$

which, when combined with Equation (33) (for $n = n_0$) gives

$$w + \frac{u_2(x_0, 0)}{n_0 u_1(x_0, 0)} = \psi_y(u_0, 0) \frac{F(n_0)}{n_0^2 f(n_0)} \left[\lambda G'(u_0) - \frac{1}{u_1(x_0, 0)} \right]. \qquad \text{...(44)}$$

Unfortunately, one cannot get much information from these " local " conditions, at least for small n. For any detail, and in particular for numerical results, one must examine the whole system of equations. It is easier to do that for particular examples of the general problem, and that is what we shall do in succeeding sections. It may be noted, however, that Equation (44) does provide us with *some* information about n_0 and x_0. For example, it is clear that n_0 can be zero only if F/nf tends to 0 as n tends to 0; indeed, since the left hand side of Equation (44) is bounded, $n_0 = 0$ only if $x_0 = 0$, and therefore $1/u_1 = 0$. It follows that $n_0 = 0$ only if $F/(n^2 f)$ is bounded as $n \to 0$, which means that F tends to zero faster than $\exp(-1/n)$. This excludes the cases usually considered by economists. We

---

[1] The analysis and result can be generalized to the utility function $u(x, z, n)$ where the parameter n can indicate variations in tastes as well as skill. The extension is fairly routine and will not be discussed here.

may conclude at this stage that it will be optimal, in the most interesting cases, to encourage some of the population to be idle.

A number of conclusions have been obtained, but they are fairly weak: the marginal tax rate lies between zero and one; in a large class of cases, consumption and labour supply vary continuously with the skill of the individual; there will usually be a group of people who ought to work only if they enjoy it. The main feature of the results is that the optimum tax schedule depends upon the distribution of skills within the population, and the labour-consumption preferences of the population, in such a complicated way that it is not possible to say in general whether marginal tax rates should be higher for high-income, low-income, or intermediate-income groups. The two integral equations that characterise the optimum tax schedule are, however, of a reasonably manageable form. One expects to be able to calculate the schedule in particular cases without great difficulty. In the next sections of the paper, we shall show how this can be done in certain special cases, and obtain further properties of the optimum tax in these cases.

6. ADDITIVE UTILITY

An interesting case arises when, for all x and y,

$$u_{12} = 0. \qquad \qquad ...(45)$$

Thus u_1 depends only on x, and u_2 only on y.

Proposition 4. *If assumption* (45) *is satisfied,* $V(x, y)$ *is an increasing function of* y, *bounded for small* x *and* y.

Proof. $V = -yu_2(y)/u_1(x)$, and $V_2 = (-u_2-yu_{22})/u_1 > 0$. Boundedness is obvious. $\parallel$

Corollary. *Under assumption* (45), *Theorem* 1 *applies.*

In particular we know, from statement (v) of that Theorem that y_n is continuous provided that ψ_y is non-decreasing. In the present case, this condition is equivalent to the requirement that

$$-yu_2(y) \text{ is convex.} \qquad \qquad ...(46)$$

There is no reason why this assumption should hold in general, but it is easily checked for any particular case. We shall now restrict attention to cases for which (46) holds.[1]

If we restrict attention also to cases where z is strictly increasing when $n > n_0$, the optimum situation will be a solution of the equations

$$\left\{ \left(w + \frac{u_2}{nu_1} \right) n^2 f(n) = \psi_y \int_n^\infty \left(\frac{1}{u_1} - \lambda G' \right) f(m) \, dm, \qquad \qquad ...(47) \right.$$

$$\left. u_n = u_0 - \int_0^n y_m u_2 \frac{dm}{m}. \qquad \qquad ...(48) \right.$$

We shall further assume that f is continuously differentiable. Since x_n, y_n are continuous in this case, it follows that u_n and $\left(w + \dfrac{u_2}{nu_1} \right) / \psi_y$ are differentiable functions of n. Write

$$v = \frac{w + \dfrac{u_2}{nu_1}}{\psi_y}. \qquad \qquad ...(49)$$

---

[1] In [4] a theorem is proved which states that the conditions of Theorem 2 are in fact *sufficient* (as well as necessary) for an optimum in the special case now being considered.

u and v are continuously differentiable functions of x and y. Since $\frac{\partial u}{\partial x} > 0$, $\frac{\partial u}{\partial y} < 0$, and,

as can easily be seen, $\frac{\partial v}{\partial x} < 0$, $\frac{\partial v}{\partial y} < 0$, the Jacobian $\frac{\partial(u, v)}{\partial(x, y)}$ is always negative. Consequently x and y can be expressed as continuously differentiable functions of u and v, and are therefore themselves differentiable functions of n.

We can now write Equations (47) and (48) as differential equations:

$$\frac{dv}{dn} = -\frac{v}{n}\left(2 + \frac{nf'}{f}\right) - \frac{1}{n^2 u_1} + \frac{\lambda G'}{n^2}, \qquad \ldots(50)$$

$$\frac{du}{dn} = -\frac{yu_2}{n}; \qquad \ldots(51)$$

which, as we have just shown, can be thought of as equations in u and v. The particular solution we seek, and the particular value of λ, are defined by the boundary conditions, Equations (39), (40),

$$v_{n_0} = \frac{F(n_0)}{n_0^2 f(n_0)}\left[\lambda G'(u_{n_0}) - \frac{1}{u_1(x_{n_0})}\right], \qquad \ldots(52)$$

which is the form (38) takes here, and

$$v_n n^2 f(n) \to 0 \quad (n \to \infty), \qquad \ldots(53)$$

which is apparent from Equation (47). Provided that z_n is strictly increasing for $n \geqq n_0$, a solution that satisfies all those conditions will, by Theorem 2 of [4], provide the optimum.

Equations (39) and (40), the production function and the marginal productivity equation, may be ignored in the calculations. Corresponding to the particular values of w and λ used in the calculation, one obtains values for X and Z. Thus we know the optimum tax schedule when the marginal product is w and the average product is X/Z. In this way one could obtain a range of tax schedules corresponding to different average products and marginal products—which is what one wants. Of course, it is desirable to choose λ so that the average product will be related to the marginal product, w, in a reasonable way. This should not present any great difficulty.

To determine the sign of $\frac{dz_n}{dn} = y_n + n\frac{dy_n}{dn}$, we calculate, from Equation (49),

$$\psi_y \frac{dv}{dn} = \left(\frac{u_{22}}{nu_1} - v\psi_{yy}\right)\frac{dy}{dn} - \frac{u_2}{n^2 u_1} - \frac{u_2 u_{11}}{nu_1^2}\frac{dx}{dn}$$

$$= \left(\frac{u_{22}}{nu_1} - v\psi_{yy}\right)\frac{dy}{dn} - \frac{u_2}{n^2 u_1} + \frac{u_2^2 u_{11}}{nu_1^3}\frac{dy}{dn} - \frac{u_2 u_{11}}{nu_1^3}\frac{du}{dn}$$

$$= \frac{1}{n}\left(\frac{u_{22}}{nu_1} - v\psi_{yy} + \frac{u_2^2 u_{11}}{nu_1^3}\right)\frac{dz}{dn} - \frac{y_n}{n}\left(\frac{u_{22}}{nu_1} - v\psi_{yy}\right) - \frac{u_2}{n^2 u_1}, \qquad \ldots(54)$$

substituting from (51). Therefore, using (50)

$$\left[\frac{u_{22}}{nu_1} - v\psi_{yy} + \frac{u_2^2 u_{11}}{nu_1^3}\right]\frac{dz}{dn} = \frac{yu_{22}}{nu_1} - yv\psi_{yy} + \frac{u_2}{nu_1} - \left(2 + \frac{nf'}{f}\right)v\psi_y - \frac{\psi_y}{nu_1} + \frac{\lambda\psi_y G'}{n}$$

$$= -\psi_y\left\{\left(2 + \frac{nf'}{f} + \frac{y\psi_{yy}}{\psi_y}\right)v + \frac{2}{nu_1} - \frac{\lambda G'}{n}\right\}. \qquad \ldots(55)$$

We may therefore check the assumption $\dfrac{dz}{dn} \geqq 0$ by examining the solution to see whether

$$\left(2 + \frac{nf'}{f} + \frac{y\psi_{yy}}{\psi_y}\right)v + \frac{2}{nu_1} - \frac{\lambda G'}{n} \geqq 0. \qquad \qquad ...(56)$$

Equation (56) is equivalent to $\dfrac{dz}{dn} \geqq 0$ because the expression in square brackets in Equation (55) is negative, term by term.

In computation, one can proceed as follows:—

[1] A value of λ is chosen. To get the right order of magnitude, one can calculate $\displaystyle\int_0^\infty u_1^{-1}fdn \Big/ \int_0^\infty G'fdn$ (cf. (38)) for some particular feasible, and *a priori* plausible, allocation of consumption and labour.

[2] A trial value of $n_0 > 0$ is chosen. (It should be borne in mind that the inequality $v_{n_0} \geqq 0$ may, with (52), restrict the range of possible n_0.)

[3] Bearing in mind that $y_{n_0} = 0$, the values of v_{n_0} and u_{n_0} are obtained from (49) and (52).

[4] The solution of equations (50) and (51) is calculated for increasing n until either (56) fails to be satisfied, or it becomes apparent that (53) will not be satisfied (see [6] below).

[5] If (56) fails to be satisfied, z_n is kept constant, u_n (and v_n) being calculated from (49) until (56) is satisfied again, when z_n is allowed to increase and the solution pursued as in [4].

[6] The attempted solution should be stopped if u_n or x_n begins to decrease, or v_n or y_n fall to zero, or x_n, y_n cannot be calculated (e.g. because u_n exceeds the upper bound of u, if there is one). Other stopping rules can be given for particular examples, depending on the structure of the solutions of the equations.

[7] A range of trial values of n_0 must be used to find the one that most nearly provides a solution satisfying (53). Efficient rules for iteration might be obtained in particular cases.

7. FEATURES OF SOLUTIONS

Solutions may, for all I know, be very diverse in their characteristics; but examination of the equations suggests a number of comments. First we note that v_n will always lie between 0 and $\dfrac{1}{\psi_y(0)}$, since

$$0 \leqq \frac{1 + \dfrac{u_2}{nu_1}}{\psi_y} \leqq \frac{1 + \dfrac{u_2}{nu_1}}{\psi_y(0)} < \frac{1}{\psi_y(0)}. \qquad \qquad ...(57)$$

We are therefore led to expect that v tends to a limit as $n \to \infty$. (It might cycle for certain forms of f, of a kind one would perhaps be unlikely to use.) y is also bounded, by 0 and 1, and is therefore likely to tend to a limit. One is then led to certain conjectures about the limits, which ought to hold for sufficiently regular f and u.

Let

$$\frac{nf'}{f} \to \gamma + 2 \leqq \infty. \qquad \qquad ...(58)$$

(Since $\int_0^\infty nf dn < \infty$, $\gamma \geqq 0$: otherwise $n^2 f$ is increasing for large n, therefore bounded below.) Further, suppose

$$u_1 \sim \alpha x^{-\mu} \quad (\mu > 0) \qquad \qquad ...(59)$$

as $x \to \infty$. Then there appear to be three cases; in each of which one expects the following results to hold.

(i) $\mu < 1$. As $n \to \infty$,

$$y_n \to 1 \qquad \qquad ...(60)$$

and $\qquad v_n \to 0.$ $\qquad \qquad ...(61)$

The marginal tax rate,

$$\theta \to 1. \qquad \qquad ...(62)$$

(ii) $\mu = 1$. As $n \to \infty$,

$$y_n \to \bar{y}, \qquad \qquad ...(63)$$

where $\bar{y}$ is defined (uniquely) by

$$\bar{y} u_2(\bar{y}) = -\alpha, \qquad \qquad ...(64)$$

and $\qquad v_n \to [-(1+\gamma)u_2(\bar{y}) - \bar{y} u_{22}(\bar{y})]^{-1}.$ $\qquad \qquad ...(65)$

Furthermore,

$$\theta \to \frac{1+v}{1+v+\gamma}, \qquad \qquad ...(66)$$

where

$$v = \frac{\bar{y} u_{22}(\bar{y})}{u_2(\bar{y})}. \qquad \qquad ...(67)$$

(iii) $\mu > 1$. As $n \to \infty$,

$$y_n \to 0, \qquad \qquad ...(68)$$

and $\qquad v_n \to [-(1+\gamma)u_2(0)]^{-1}.$ $\qquad \qquad ...(69)$

$$\theta \to \frac{1}{1+\gamma}. \qquad \qquad ...(70)$$

(It may be noted that, in a natural sense, (66) holds for all cases.)

Before indicating the reasons for these conjectures, a few words of interpretation may be in place. On the whole, the distribution of income from employment appears to be of Paretian form at the upper tail[1]: Equation (58) holds with γ between 1 and 2, roughly speaking. It is not improbable, however, that marginal productivity per working year is distributed differently from actual incomes: the lognormal distribution is the most plausible simple distribution. For this, $\gamma = \infty$, and

$$-\frac{nf'}{f} \sim \frac{\log n}{\sigma^2} \qquad \qquad ...(71)$$

for large n; (σ^2 is the variance of the distribution of logarithm of incomes).

---

[1] See the general assessment by Lydall [3].

The realism of alternative assumptions about utility may be assessed by calculating the response of the consumer to a linear budget constraint, $x = wy + a$. It is easy to see that utility-maximization requires (since $u_{12} \equiv 0$)

$$- \frac{u_1(x)}{u_2(y)} = \frac{1}{w}, \quad x = wy + a. \qquad \ldots(72)$$

If $u_1 = \alpha x^{-\mu}$, we have to solve

$$\alpha w = -(a + wy)^{\mu} u_2(y). \qquad \ldots(73)$$

(If $\alpha w \leqq -a^{\mu} u_2(0)$, $y = 0$.)   Clearly the solution has the following properties:

$$\left. \begin{array}{l} y \to 1 \text{ as } w \to \infty \text{ if } \mu < 1, \\ y \to 0 \text{ as } w \to \infty \text{ if } \mu > 1. \end{array} \right\} \qquad \ldots(74)$$

(Cf. (61) and (68).)   Also

$$\left. \begin{array}{ll} x \sim a + w & (\mu < 1), \\ x \sim \left(\dfrac{aw}{u_2(0)} \right)^{\frac{1}{\mu}} & (\mu > 1) \end{array} \right\}. \qquad \ldots(75)$$

These asymptotic properties suggest that the case $\mu = 1$ is particularly interesting.
When $\mu = 1$, since, by (73)

$$\frac{a}{w} = - \frac{\alpha}{u_2} - y,$$

$$- y u_2 \to \alpha \text{ as } w \to \infty;$$

i.e.

$$y \to \bar{y}, \qquad \ldots(76)$$

where $\bar{y}$ is defined by (64).   (Cf. (63).)   If in addition,

$$u_2(y) = -(1 - y)^{-\delta} \quad (\delta > 0), \qquad \ldots(77)$$

we have

$$\bar{y}(1 - \bar{y})^{-\delta} = \alpha,$$

$$\bar{y}(1 - \bar{y})^{-1} = v.$$

The choice of α may be influenced by considering that $y = 0$ when $w/a \leqq 1/\alpha$.   It is interesting to note that, if

$$\alpha = 2, \quad \delta = 1, \quad \gamma = 2,$$

$$\bar{y} = 2/3, \quad v = 2,$$

and, if our conjectures are correct,

$$\theta \to 60 \text{ per cent.}$$

This case is perhaps not completely unrealistic; but it should be remembered that the homogeneous form for u means that the decision not to work depends only on the *ratio* of earned to unearned income, which is not a very realistic assumption.

It will be noticed that, in this case, the asymptotic marginal tax rate is very sensitive to the value of μ (in the neighbourhood of 1).

The reasons for the conjectures Equations (60)-(70) (in fact, I can provide a proof of (iii) and will do so below) are as follows.   One expects that, as $n \to \infty$, the relevant solution of the differential equations will tend towards a singularity of the equations: not only will y and v tend to limits, but $n \dfrac{dy}{dn}$ and $n \dfrac{dv}{dn}$ will tend to zero.   Denote the postulated limit of y_n by $\bar{y}$.   Consider first the case $u_1 = \alpha x^{-\mu} (\mu < 1)$.

In this case utility is unbounded. I shall show that $\bar{y} = 1$. If not, u_2 and ψ_y tend to finite limits, and, from (51), we have

$$nu_1 \frac{dx}{dn} = -u_2\left(y + n\frac{dy}{dn}\right) \to -\bar{y}u_2(\bar{y}). \qquad \text{...(78)}$$

Therefore, since $u_1 \dfrac{dx}{dn} = \alpha \dfrac{d}{dn}\left[\dfrac{1}{1-\mu} x^{1-\mu}\right]$,

$$\frac{\alpha}{1-\mu} x^{1-\mu} = -\bar{y}u_2(\bar{y})\log n[1+o(1)]. \qquad \text{...(79)}$$

This implies that

$$nu_1 = 0\left[n(\log n)^{-\frac{\mu}{1-\mu}}\right] \qquad \text{...(80)}$$

$$\to \infty. \qquad \text{...(81)}$$

Therefore

$$\frac{1}{n^2 f(n)}\int_n^\infty \left[\frac{1}{u_1} - \lambda\right]f(m)dm = \frac{1}{\psi_y}\left[1 + \frac{u_2}{nu_1}\right] \to \frac{1}{\psi_y(\bar{y})} > 0, \qquad \text{...(82)}$$

which is readily seen to be inconsistent with (80) if the distribution is either Paretian or lognormal.

We must therefore expect that $\bar{y} = 1$. Suppose now that $1 + \dfrac{u_2}{nu_1}$, the marginal tax rate, tends to a limit $\bar{\imath} < 1$. Then

$$\frac{dx}{dn} = -\frac{u_2}{nu_1}\left(y + n\frac{dy}{dn}\right) \to 1 + \bar{\imath}, \qquad \text{...(83)}$$

and consequently

$$\frac{x}{n} \to 1 - \bar{\imath}. \qquad \text{...(84)}$$

This implies that

$$\frac{1}{u_1} = \frac{1}{\alpha}(1-\bar{\imath})^\mu n^\mu[1+o(1)], \qquad \text{...(85)}$$

from which we can deduce the behaviour of

$$I = \frac{u_1}{nf}\int_n^\infty \left(\frac{1}{u_1} - \lambda\right)f(m)dm \qquad \text{...(86)}$$

as $n \to \infty$. In the Paretian case, $f \sim n^{-2-\gamma}$, it is easily seen that

$$I \to (2 + \gamma - \mu)^{-1} > 0. \qquad \text{...(87)}$$

Since $1 - \bar{\imath} = \lim \dfrac{\psi_y}{u_2}.\dfrac{u_2}{nu_1}.I$, and $\dfrac{\psi_y}{u_2} = -1 - \dfrac{d}{dy}\log|u_2|$ tends to $-\infty$ as $y \to 1$ (if it tends to a limit at all), we must have $\dfrac{u_2}{nu_1} \to 0$, which is inconsistent with the assumption $\bar{\imath} < 1$. In the lognormal case, one obtains

$$1 - \bar{\imath} = \lim \frac{\psi_y}{u_2}.\frac{u_2}{nu_1}\frac{\text{constant}}{\log n}. \qquad \text{...(88)}$$

If $\dfrac{\psi_y}{u_2}\dfrac{1}{\log n}$ tended to a finite limit, since

$$\log|u_2| \sim \log(1-\bar{\imath}) + \log(nu_1) \sim (1-\mu)\log n,$$

$\dfrac{1}{\log|u_2|}\dfrac{d}{dy}\log|u_2|$ would tend to a finite limit as $y\to1$; which is clearly impossible.
Thus in the lognormal case too, we expect that $\bar{\imath}=1$. This explains the conjectures in the case $\mu<1$.

If $\mu=1$,

$$\alpha\frac{d(\log x)}{d(\log n)}=nu_1\frac{dx}{dn}=-u_2\left(y+n\frac{dy}{dn}\right),\qquad\qquad\ldots(89)$$

which therefore cannot tend to ∞, since in that case $u_1^{-1}=\dfrac{x}{\alpha}>n^M$ eventually for any

finite M, so that $\dfrac{1}{n^2f(n)}\displaystyle\int_n^\infty\dfrac{1}{u_1}f(m)dm$ becomes unbounded as $n\to\infty$.

We can expect, therefore, that $y\to\bar{y}<1$ and

$$\alpha\frac{\log x}{\log n}\to-\bar{y}u_2(\bar{y}).\qquad\qquad\ldots(90)$$

It is easily seen that the only plausible value of $\bar{y}$ is that for which $\log x/\log n\to1$, i.e.

$$\bar{y}u_2(\bar{y})=-\alpha.\qquad\qquad\ldots(91)$$

Then if $1+\dfrac{u_2}{nu_1}\to\bar{\imath}$, we shall have

$$n\alpha x^{-1}\to\frac{-u_2(\bar{y})}{1-\bar{\imath}},$$

and

$$\frac{\psi_y}{n^2f(n)}\int\left(\frac{1}{u_1}-\lambda\right)f(m)dm\to\frac{(1-\bar{\imath})\psi_y(\bar{y})}{-u_2(\bar{y})}\frac{1}{\gamma};$$

which suggests that

$$\bar{\imath}=(1-\bar{\imath})\frac{\psi_y(\bar{y})}{-u_2(\bar{y})}\frac{1}{\gamma}$$

$$=(1-\bar{\imath})(1+v)/\gamma,\qquad\qquad\ldots(92)$$

in the notation (57). This is equivalent to (56). In particular, we expect that $\bar{\imath}=0$ in the lognormal case.

When $\mu>1$, the utility function is bounded above, and a more general and rigorous treatment is easy. u_n is an increasing function, and being now bounded tends to a finite $\bar{u}$. We shall write

$$u(x,y)=\chi(x)+\rho(y).\qquad\qquad\ldots(93)$$

Since x is an increasing function, $\chi(x)$ also tends to a finite limit $\bar{\chi}$. Thus $\rho(y)$ tends to a limit, and so does y. The limit of y must be zero, since otherwise (32) implies $u\to\infty$, which is now impossible.

Now

$$v+\frac{1}{nu_1}=\frac{1}{\psi_y}+\left(1+\frac{u_2}{\psi_y}\right)\frac{1}{nu_1}$$

$$\to\frac{1}{-u_2(0)}\qquad\qquad\ldots(94)$$

in this case $\left(\text{since } \dfrac{1}{nu_1}, \text{ being} \leqq \dfrac{1}{-u_2}, \text{ is bounded}\right)$.   Therefore Equation (50) becomes

$$n\frac{dv}{dn} = (\gamma+1+o(1))v + \frac{1}{u_2(0)} + o(1) \qquad \ldots(95)$$

in the Paretian case.   From (95) one deduces, by the usual method of solving a first-order linear differential equation, that

$$v \to \frac{1}{-u_2(0)(\gamma+1)}, \qquad \ldots(96)$$

from which it follows at once that the marginal tax rate tends to $(\gamma+1)^{-1}$.   It is easily checked that in the lognormal case the marginal tax rate tends to zero.

In the next section, a particular case is examined in detail, and provides confirmation for some of our conjectures.

8. AN EXAMPLE

Case I.   Let us, by way of illustration, analyze the following case:

$$u = \alpha \log x + \log(1-y)$$

$$G(u) = -\frac{1}{\beta} e^{-\beta u} \quad (\beta \geqq 0) \ ^1 \qquad \ldots(97)$$

$$f(n) = \frac{1}{n} \exp\left[-\frac{(\log n+1)^2}{2}\right].$$

(The last assumes a lognormal distribution of skills: the average of n is $\dfrac{1}{\sqrt{e}} = 0\cdot607\ldots$).

We put $w = 1$.   With these assumptions, Equations (50) and (51) become

$$\frac{dv}{dn} = v\frac{\log n}{n} - \frac{x}{\alpha n^2} + \frac{\lambda}{n^2} e^{-\beta u},$$

$$\frac{du}{dn} = \frac{y}{n(1-y)},$$

where

$$v = \left[1+\frac{u_2}{nu_1}\right]/\psi_y = \frac{1-\dfrac{x}{\alpha n(1-y)}}{1/(1-y)^2} = (1-y)\left(1-y-\frac{x}{\alpha n}\right),$$

and

$$e^u = x^\alpha(1-y).$$

For simplicity, we consider the case $\beta = 0$ first, and put

$$s = 1-y,$$
$$t = \log n.$$

The equations become, since $u = \alpha \log(\alpha n) + \alpha \log\left(s-\dfrac{v}{s}\right) + \log s$,

$$\frac{dv}{dt} = v\left(t+\frac{1}{s}\right) - s + \lambda e^{-t}, \qquad \ldots(98)$$

$$\frac{ds}{dt} = \frac{[1-\alpha-(1+\alpha)s](s^2-v)+\alpha s(vt+\lambda e^{-t})}{(1+\alpha)s^2-(1-\alpha)v}. \qquad \ldots(99)$$

---

[1] In the case of $\beta = 0$, we define $G = u$.

Solutions of these equations are depicted in Fig. 1. We now establish their properties. We remember that, in the optimum solution, $0 < v < s^2$ (for the marginal tax rate, v/s^2, is between 0 and 1). Using this fact, we can deduce from the first equation that

$$v \to 0 \quad (t \to \infty).$$

Suppose that, for some t, $vt \geq 1$. Then

$$\frac{d}{dt} v \geq vt + \frac{v - s^2}{s} > vt - 1 \geq 0,$$
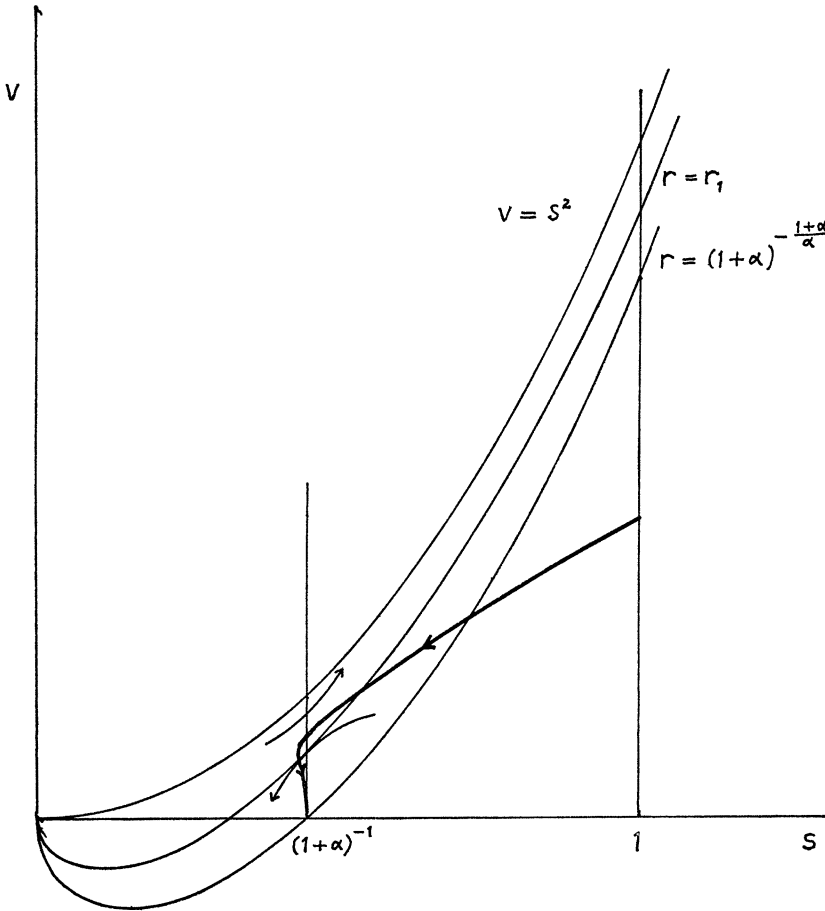
FIGURE 1

since $v > 0$, and $s \leq 1$. Therefore v is increasing at an increasing rate, contradicting $v < s^2 \leq 1$. This shows that, in fact,

$$0 < v < 1/t. \qquad \qquad \qquad ...(100)$$

The two equations together imply that

$$\frac{d}{dt} \left[s^{\frac{1-\alpha}{\alpha}}(s^2 - v) \right] = \frac{1 - (1+\alpha)s}{s} s^{\frac{1-\alpha}{\alpha}}(s^2 - v), \qquad \qquad ...(101)$$

as one may see if one multiplies the first by αs, and the second by $[(1+\alpha)s^2 - (1-\alpha)v]$, and subtracts. Write

$$r = s^{\frac{1-\alpha}{\alpha}}(s^2 - v). \qquad \qquad \qquad ...(102)$$

so that

$$\alpha \frac{dr}{dt} = \frac{1-(1+\alpha)s}{s}\, r. \qquad \text{...(103)}$$

When $s < \dfrac{1}{1+\alpha}$, r increases; when $s > \dfrac{1}{1+\alpha}$, r decreases. For this reason s cannot tend to a limit other than $1/(1+\alpha)$: we shall show more, that $s \to 1/(1+\alpha)$. (Cf. Fig. 1.)

Since $v \to 0$, given $\varepsilon > 0$, there exists t_0 such that $0 < v_t < \varepsilon$ for all $t \geqq t_0$. Then

$$s^{\frac{1+\alpha}{\alpha}} - \varepsilon s^{\frac{1-\alpha}{\alpha}} < r < s^{\frac{1+\alpha}{\alpha}} \quad (t \geqq t_0). \qquad \text{...(104)}$$

If $r_t > (1+\alpha)^{-\frac{1+\alpha}{\alpha}}$, the right hand inequality implies that

$$s_t > \frac{1}{1+\alpha}. \qquad \text{...(105)}$$

Therefore r is decreasing. If

$$r_t < (1+\alpha)^{-\frac{1+\alpha}{\alpha}} - \varepsilon \max\left[1, (1+\alpha)^{-\frac{1-\alpha}{\alpha}}\right], \qquad \text{...(106)}$$

we obtain from the left hand inequality (104),

$$s_t^{\frac{1+\alpha}{\alpha}} < (1+\alpha)^{-\frac{1+\alpha}{\alpha}} - \varepsilon\left\{\max\left[1, (1+\alpha)^{-\frac{1-\alpha}{\alpha}}\right] - s_t^{\frac{1-\alpha}{\alpha}}\right\} \leqq (1+\alpha)^{-\frac{1+\alpha}{\alpha}} \qquad \text{...(107)}$$

if, either $\alpha \leqq 1$ (in which case $\{...\} \geqq 0$ since $s \leqq 1$), or $\alpha > 1$ and $s_t \geqq \dfrac{1}{1+\alpha}$. Thus, in fact

$$s_t < \frac{1}{1+\alpha}, \qquad \text{...(108)}$$

and, by (98), r_t is increasing. Combining these two results, we deduce that

$$r_t \to (1+\alpha)^{-\frac{1+\alpha}{\alpha}},$$

which in turn implies, since $v > 0$, that

$$s_t \to \frac{1}{1+\alpha}. \qquad \text{...(109)}$$

Our demonstration that v and s tend to limits 0 and $\dfrac{1}{1+\alpha}$, respectively, confirms the conjectures for the special case. It is readily checked that exactly the same arguments apply to the case $\beta > 0$. As we have noted previously, the marginal tax rate is v/s^2. Thus, as $t \to \infty$

$$\theta \to 0. \qquad \text{...(110)}$$

It is a striking result; but we should note at once that 0 is a poor approximation to v/s^2 even for large t. This becomes apparent when we demonstrate that $vt \to \dfrac{1}{1+\alpha}$.

Suppose the contrary, that $\left| vt - \dfrac{1}{1+\alpha} \right| > \varepsilon > 0$ for an unbounded set of values of t.

If $vt > \dfrac{1}{1+\alpha} + \varepsilon$, and t is large enough to imply that $s_t < \dfrac{1}{1+\alpha} + \frac{1}{2}\varepsilon$,

$$\frac{dv}{dt} > \frac{1}{2}\varepsilon. \qquad \text{...(111)}$$

Thus vt continues greater than $\dfrac{1}{1+\alpha}+\varepsilon$, and $\dfrac{dv}{dt}>\tfrac{1}{2}\varepsilon$ for all larger t: but this implies that

$v\to\infty$, which we have already shown to be false. If on the other hand $vt<\dfrac{1}{1+\alpha}-\varepsilon$, and

t is greater than $\dfrac{2}{\varepsilon}$, and is large enough to imply

$$v_t<\frac{1+\alpha}{4}, \quad \lambda te^{-t}<\frac{1+\alpha}{4}, \quad s_t>\frac{1}{1+\alpha}-\tfrac{1}{2}\varepsilon,$$

then

$$\frac{d}{dt}(vt)=v+t(vt-s)+\frac{vt}{s}+\lambda te^{-t}$$

$$<\frac{1+\alpha}{2}+\frac{\dfrac{1}{1+\alpha}-\varepsilon}{\dfrac{1}{1+\alpha}-\tfrac{1}{2}\varepsilon}-\tfrac{1}{2}\varepsilon t$$

$$<1-\tfrac{3}{2}=-\tfrac{1}{2}. \qquad\qquad \text{...(112)}$$

This implies that vt becomes negative, which is impossible. Therefore $\left|vt-\dfrac{1}{1+\alpha}\right|<\varepsilon$ for all large enough t:

$$vt\to\frac{1}{1+\alpha}. \qquad\qquad \text{...(113)}$$

Thus

$$\theta=v/s^2\sim\frac{1+\alpha}{t}. \qquad\qquad \text{...(114)}$$

Only 1 per cent of our population have $t\geqq 1\cdot 7$ (one in a thousand have $t\geqq 2\cdot 4$). Since one might want to have α as low as 1, the above approximation is clearly rather bad even at $t=2$.[1,2] How bad will become apparent in the next section.

Case II. It is also of interest to examine the case of a skill-distribution with Paretian tail:

$$\frac{nf'}{f}\to\gamma+2, \quad \gamma>0. \qquad\qquad \text{...(115)}$$

The equations for the optimum become (with $\beta=0$),

$$\frac{dv}{dt}=v\gamma(t)+\frac{v}{s}-s+\lambda e^{-t}, \qquad\qquad \text{...(116)}$$

$$\frac{ds}{dt}=\frac{[1-\alpha-(1+\alpha)s](s^2-v)+\alpha s(v\gamma(t)+\lambda e^{-t})}{(1+\alpha)s^2-(1-\alpha)v}; \qquad\qquad \text{...(117)}$$

[1] In this example, $\sigma^2=1$: that is, the standard deviation of log n is 1. This is done merely for convenience in manipulations. A precisely similar theory holds for a general lognormal distribution.

It can be shown, by continuing the methods of the text, that $vt\sim\dfrac{1}{1+\alpha}-\dfrac{1}{t}$ while $s=\dfrac{1}{1+\alpha}+o\left(\dfrac{1}{t^2}\right)$.

The fact that the optimum path is tangential to the vertical at $(s,v)=\left(\dfrac{1}{1+\alpha},0\right)$ implies that $s<\dfrac{1}{1+\alpha}$, for large t, since otherwise r would be decreasing, and that, as can be seen from the diagram, is inconsistent with $\dfrac{dv}{ds}\to\infty$. Thus we have the situation portrayed in Fig. 1.

[2] The case $\beta>0$ can be treated in a precisely similar way, to obtain the same qualitative results.

and, exactly as before, one has the equation

$$\frac{dr}{dt} = \frac{1-(1+\alpha)s}{s}r \qquad \qquad ...(118)$$

where $r = s^{\frac{1+\alpha}{\alpha}}(s^2-v)$. The situation is portrayed in Fig. 2. The broken curves have equations

$$s^{\frac{1-\alpha}{\alpha}}(s^2-v) = r_i \quad (i = 1, 2, 3) \qquad \qquad ...(119)$$
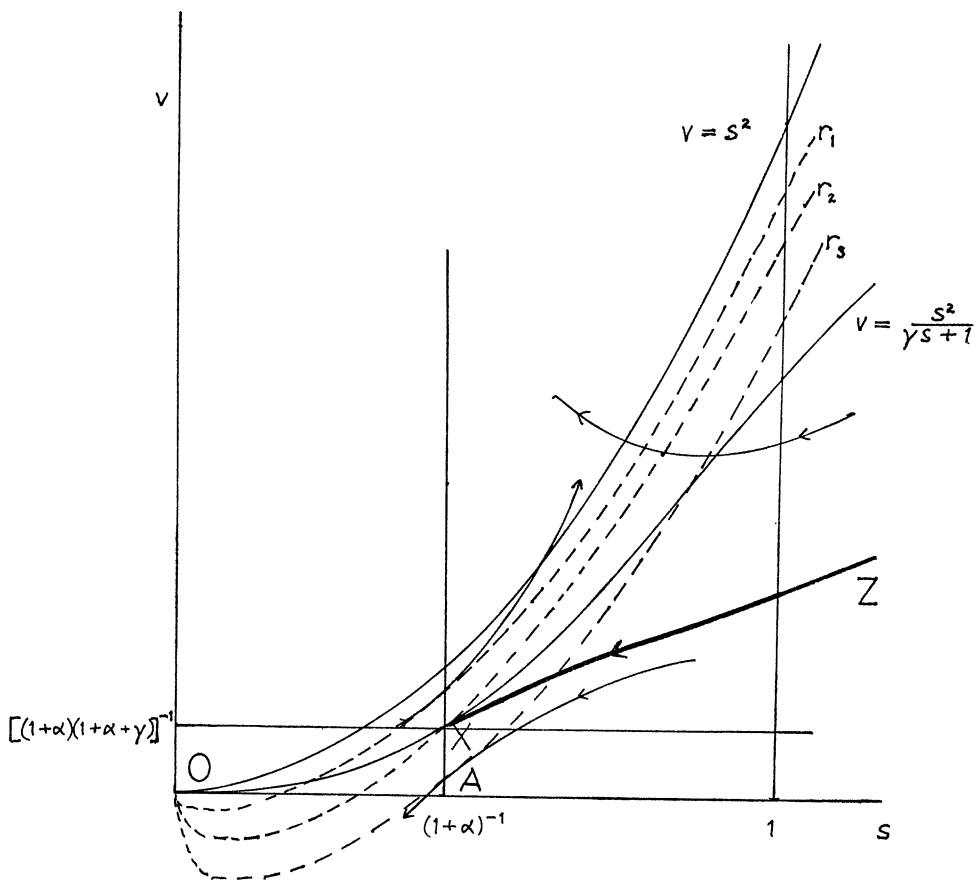
FIGURE 2

with $0 < r_1 < r_2 < r_3$. It will be noted that such a curve, with equation

$$v_1 = s^2 - rs^{\frac{1-\alpha}{\alpha}} \quad (r \text{ constant}), \qquad \qquad ...(120)$$

always cuts from below the curve

$$v_2 = \frac{s^2}{\gamma s + 1} + p \quad (p \text{ constant}) \qquad \qquad ...(121)$$

that passes through the same point. This follows from the calculation,

$$\frac{dv_1}{ds} - \frac{dv_2}{ds} = \frac{d}{ds}\left(\frac{\gamma s^3}{\gamma s + 1} - rs^{\frac{1-\alpha}{\alpha}}\right) > 0. \qquad \qquad ...(122)$$

N

This remark will prove very useful; but first we want to establish that, for large t, the sign of $\dfrac{dv}{dt}$ is nearly the same as the sign of $v - \dfrac{s^2}{\gamma s + 1}$.

Let ε' be a positive number, and let t_1 be so large that $|v\gamma(t) + \lambda e^{-t} - v\gamma| > \varepsilon'$ when $t \geq t_1$. Since $s = 1$ at $t_0 = \log n_0$, $s < \dfrac{1}{1+\alpha}$ at t only if $s = \dfrac{1}{1+\alpha}$ for some previous t_1; if (for the given t) t_1 is the greatest such, we have from Equation (118)

$$r_t > r_{t_1} = (1+\alpha)^{-\frac{1-\alpha}{\alpha}} \left(\frac{1}{(1+\alpha)^2} - v_{t_1} \right)$$

$$\geq (1-\alpha)^{-\frac{1-\alpha}{\alpha}} \inf \left\{ \frac{1}{(1+\alpha)^2} - v_t \, \middle| \, s_t = \frac{1}{1+\alpha}, \frac{dv_t}{dt} < 0 \right\}$$

$$= \Delta > 0, \qquad\qquad \text{...(123)}$$

since as $t \to \infty$, $0 > \dfrac{dv}{dt}$ implies

$$v_t < \frac{s_t^2}{\gamma s_t + 1} + o(1)$$

$$< s_t^2 - \gamma s_t^3 + o(1). \qquad\qquad \text{...(124)}$$

Therefore s_t is positively bounded below, say

$$s_t \geq \Delta' > 0. \qquad\qquad \text{...(125)}$$

Hence, when $t \geq t_1$,

$$\frac{dv}{dt} = v\gamma(t) + \frac{v}{s} - s + \lambda e^{-t}$$

$$> \left(v - \frac{s^2}{\gamma s + 1} \right)\left(\gamma + \frac{1}{s} \right) - \varepsilon'$$

$$> \varepsilon', \qquad\qquad \text{...(126)}$$

if

$$v > \frac{s^2}{\gamma s + 1} + \frac{2\varepsilon'}{\gamma + \dfrac{1}{\Delta'}}. \qquad\qquad \text{...(127)}$$

Similarly, we can show that

$$\frac{dv}{dt} < -\varepsilon' \qquad\qquad \text{...(128)}$$

if

$$v < \frac{s^2}{\gamma s + 1} - \frac{2\varepsilon'}{\gamma + \dfrac{1}{\Delta'}}. \qquad\qquad \text{...(129)}$$

Now write $\varepsilon = \varepsilon' / \left(\gamma + \dfrac{1}{\Delta'} \right)$. It is clear that, if, for some $t \geq t_1$,

$$v > \frac{s^2}{\gamma s + 1} + \varepsilon \quad \text{and} \quad s \geq \frac{1}{1+\alpha},$$

then $\dfrac{dv}{dt} > 0$ and also $\dfrac{dr}{dt} < 0$. Therefore, by the properties of the two sets of curves (cf.

Fig. 3), $v - \dfrac{s^2}{\gamma s + 1}$ is increasing. Thus for all subsequent t, $\dfrac{dv}{dt} > \varepsilon'$, and $v \to \infty$. Such a path cannot be optimum. Consequently on the optimum path, if $t \geqq t_1$,

$$\text{either } s < \frac{1}{1+\alpha} \text{ or } v \leqq \frac{s^2}{\gamma s + 1} + \varepsilon. \qquad \ldots(130)$$

Similarly, for $t \geqq t_1$,

$$\text{either } s > \frac{1}{1+\alpha} \text{ or } v \geqq \frac{s^2}{\gamma s + 1} - \varepsilon. \qquad \ldots(131)$$
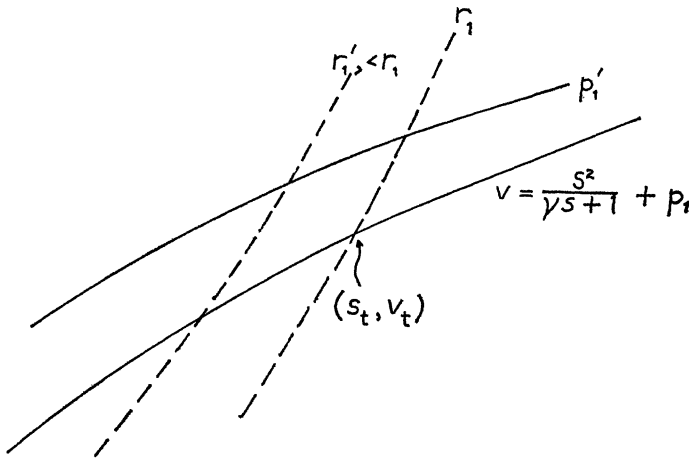
FIGURE 3

Suppose that at t_1, $s > \dfrac{1}{1+\alpha}$. $\left(\text{An exactly similar argument applies if } s < \dfrac{1}{1+\alpha}.\right)$ Then r is decreasing, and continues to do so until

$$r = r' = (1+\alpha)^{-\frac{1-\alpha}{\alpha}} \left(\frac{1}{(1+\alpha)^2} \frac{\gamma}{(1+\alpha+\gamma)} + \varepsilon \right).$$

Only then can s become less than $\dfrac{1}{1+\alpha}$. (Cf. Fig. 4.) Once $s < \dfrac{1}{1+\alpha}$, r increases again. Therefore at no time is

$$r < r'' = (1+\alpha)^{-\frac{1-\alpha}{\alpha}} \left(\frac{\gamma}{(1+\alpha)^2(1+\alpha+\gamma)} - \varepsilon \right).$$

Nor can we have $r > r'$ at any later time. Thus we have found t_2 such that, when $t \geqq t_2$, (s_t, v_t) lies in the curvilinear parallelogram $LMPQ$ in Fig. 4, which contains X, and can be made as small as we please by suitable choice of ε'. Therefore as $t \to \infty$,

$$s_t \to \frac{1}{1+\alpha}, \quad v_t \to \frac{1}{(1+\alpha)(1+\alpha+\gamma)}. \qquad \ldots(132)$$

The optimum path is indicated by XZ in Fig. 2. On it, the marginal tax rate,

$$\theta = \frac{v_t}{s_t^2} \to \frac{1+\alpha}{1+\alpha+\gamma}; \qquad\qquad \dots(133)$$

which confirms our conjecture in this special case.[1,2]

It should be noted that we have not shown, in either of these cases, that s diminishes (nor even that $z = ny = e^t(1-s)$ increases) all along the path: the possibility that z is constant for some range of n, in the optimum regime, remains in both the examples we have discussed. Calculation of specific cases is required to settle this issue. Such calculation is not difficult with the information about the solution that we now have.
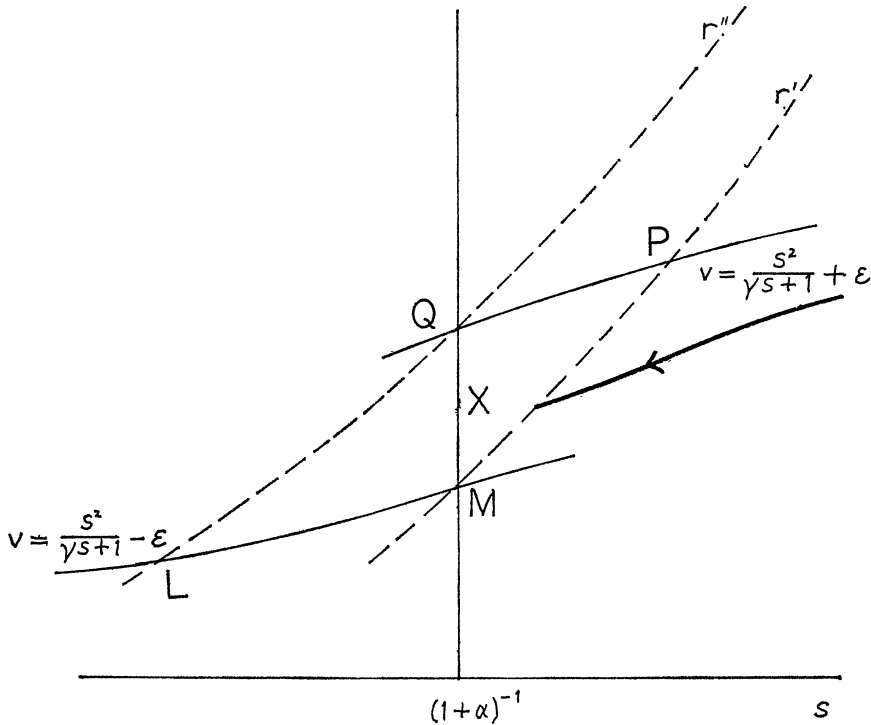
FIGURE 4

9. A NUMERICAL ILLUSTRATION

The computations whose results are presented in the tables below were carried out for the first case examined above, with $\alpha = 1$, but with a more realistic value for σ^2. Computations have also been carried out for the case $\sigma^2 = 1$, and these provide an interesting contrast to the main set of calculations. In all cases, we take $w = 1$; and for computational convenience, the average of $\log n$ is -1. This means that the average marginal product of a full day's work is $e^{\frac{1}{2}\sigma^2 - 1}$, but it amounts only to a choice of units for the consumption good. The results show, for particular values of the average product of labour, X/Z, what is the optimum tax schedule, and what is the distribution of consumption and labour in the population.

---

[1] The case $\beta > 0$ can be treated in a precisely similar way, to obtain the same qualitative results.

[2] It is possible to calculate optimum tax schedules explicitly for a uniform (rectangular) distribution of skills; but since that distribution is of no great interest in the present context, the analysis is omitted.

For purposes of comparison, one naturally wants to know what would have been the optimum position if it had been possible to use lump-sum taxation (or, equivalently, direction of labour). Let us consider this first for the case $\beta = 0$. We shall assume a linear production function

$$X = Z + a \qquad \qquad \ldots(134)$$

(which one thinks of as applying only over a certain range of values of Z, including all those that are to be considered). In the full optimum, we maximize

$$\int [\log x + \log (1-y)] f(n) dn$$

subject to $\qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \ldots(135)$

$$\int x f(n) dn = \int n y f(n) dn + a.$$

It is clear that x will be the same for everyone:

$$x = x^0, \qquad \qquad \ldots(136)$$

and that y_n must maximize

$$\log (1-y) + ny/x^0, \qquad \qquad \ldots(137)$$

for otherwise we could improve matters by changing y_n (for a set of n of positive measure, of course) and changing the constant x correspondingly. Maximization of (137) yields

$$y_n^0 = [1 - x^0/n]_+, \qquad \qquad \ldots(138)$$

where the notation $[\ldots]_+$ means max $(0, \ldots)$.

It is worth noticing that in the full optimum, only men for whom $n > x^0$ actually work, and an interesting curiosity that, with the particular welfare function specified in (135), utility will be less for more highly skilled individuals. This is, as we have seen, impossible under the income-tax. The value of x^0 is determined by the production constraint:

$$x^0 = \int_{x^0}^{\infty} (n - x^0) f(n) dn + a, \qquad \qquad \ldots(139)$$

where, for convenience, we have taken $\int_0^{\infty} f(n) dn = 1$. In the case of the special lognormal distribution used here, it can be shown that this equation reduces to

$$2x^0 - x^0 F(x^0) - e^{\frac{1}{2}\sigma^2 - 1}[1 - F(e^{-\sigma^2} x^0)] = a. \qquad \qquad \ldots(140)$$

Solution of this equation gives the consumption level in the full optimum, and also the skill-level below which no work is required of a man, namely that at which a full day's labour would provide a wage equal to the consumption level.

When $\beta > 0$, a similar theory holds. In that case, $x > x^0$ for men with $n > x^0$, but it is still the case that such men are made to have a lower utility level than their less skilled neighbours. The equation corresponding to (140) is a little more complicated and will not be reproduced. For $n > x^0$, consumption and labour are

$$x_n = (x^0)^{(1+\beta)/(1+2\beta)} n^{\beta/(1+2\beta)},$$
$$\qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \ldots(141)$$
$$y_n = 1 - (x^0/n)^{(1+\beta)/(1+2\beta)}.$$

In the tables, certain features of the optimal regime under income taxation are given, along with x^0 for the full optimum for the same linear production function. In Tables I–X, the lognormal distribution has parameters $\sigma = 0.39$. This figure is derived from Lydall's figures for the distribution of income from employment for various countries ([3], p. 153). It is intended to represent a realistic distribution of skills within the population. In each

TABLE I

(*Case* 1)

$\alpha = 1$, $\beta = 0$, $\sigma = 0.39$, mean $n = 0.40$, $X/Z = 0.93$.
Full optimum for $X = Z - 0.013$: $x^0 = 0.19$, $F(x^0) = 0.045$.
Partial optimum (income-tax): $x_0 = 0.03$, $n_0 = 0.04$, $F(n_0) = 0.000$.

$F(n)$	x	y	$x(1-y)$	z	full optimum x
0	0·03	0	0·03	0	0·19
0·10	0·10	0·42	0·05	0·09	0·19
0·50	0·16	0·45	0·08	0·17	0·19
0·90	0·25	0·48	0·13	0·29	0·19
0·99	0·38	0·49	0·19	0·45	0·19
Population average	0·17			0·18	0·19

TABLE II

Same case as Table I.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0·03		23
0·05	0·07	−34	26
0·10	0·10	−5	24
0·20	0·18	9	21
0·30	0·26	13	19
0·40	0·34	14	18
0·50	0·43	15	16

TABLE III

(*Case* 2)

$\alpha = 1$, $\beta = 0$, $\sigma = 0.39$, mean $n = 0.40$, $X/Z = 1.10$.
Full optimum for $X = Z + 0.017$: $x^0 = 0.21$, $F(x^0) = 0.075$.
Partial optimum (income-tax): $x_0 = 0.05$, $n_0 = 0.06$, $F(n_0) = 0.000$.

$F(n)$	x	y	$x(1-y)$	z	Full optimum x
0	0·05	0	0·05	0	0·21
0·10	0·11	0·36	0·07	0·08	0·21
0·50	0·17	0·42	0·10	0·15	0·21
0·90	0·27	0·45	0·15	0·28	0·21
0·99	0·40	0·47	0·21	0·43	0·21
Population average	0·18			0·17	0·21

TABLE IV

Same case as Table III.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0·05		
0·05	0·09	−80	21
0·10	0·13	−30	20
0·20	0·21	−5	19
0·30	0·29	3	17
0·40	0·37	6	16
0·50	0·46	8	15

TABLE V
(Case 3)
$\alpha = 1, \beta = 1, \sigma = 0.39$, mean $n = 0.40$, $X/Z = 1.20$.
Full optimum for $X = Z + 0.030$: $x^0 = 0.16$, $F(x^0) = 0.016$.
Partial optimum (income-tax): $x_0 = 0.07$, $n_0 = 0.09$, $F(n_0) = 0.000$.

$F(n)$	x	y	$x(1-y)$	z	Full optimum x
0	0.07	0	0.07	0	0.16
0.10	0.12	0.28	0.08	0.07	0.18
0.50	0.17	0.37	0.11	0.14	0.21
0.90	0.26	0.43	0.15	0.26	0.25
0.99	0.39	0.46	0.21	0.42	0.29
Population average	0.18			0.15	0.21

TABLE VI
Same case as Table V.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0.07		23
0.05	0.11	−113	28
0.10	0.14	−42	27
0.20	0.22	−8	25
0.30	0.29	2	23
0.40	0.37	7	21
0.50	0.45	10	19

TABLE VII
(Case 4)
$\alpha = 1, \beta = 1, \sigma = 0.39$, mean $n = 0.40$, $X/Z = 0.98$.
Full optimum for $X = Z - 0.003$: $x^0 = 0.14$, $F(x^0) = 0.007$.
Partial optimum (income-tax): $x_0 = 0.05$, $n_0 = 0.07$, $F(n_0) = 0.000$.

$F(n)$	x	y	$x(1-y)$	z	Full optimum x
0	0.05	0	0.05	0	0.14
0.10	0.10	0.33	0.07	0.08	0.17
0.50	0.15	0.41	0.09	0.15	0.20
0.90	0.24	0.46	0.13	0.28	0.23
0.99	0.37	0.48	0.19	0.44	0.26
Population average	0.16			0.17	0.19

TABLE VIII
Same case as Table VII.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0.05		30
0.05	0.08	−66	34
0.10	0.12	−34	32
0.20	0.19	7	28
0.30	0.26	13	25
0.40	0.34	16	22
0.50	0.41	17	20

TABLE IX
(Case 5)
$\alpha = 1$, $\beta = 1$, $\sigma = 0.39$, mean $n = 0.40$, $X/Z = 0.88$.
Full optimum for $X = Z - 0.021$; $x^0 = 0.13$, $F(x^0) = 0.004$.
Partial optimum (income-tax): $x_0 = 0.04$, $n_0 = 0.06$, $F(n_0) = 0.000$.

$F(n)$	x	y	$x(1-y)$	z	Full optimum x
0	0·04	0	0·04	0	0·13
0·10	0·09	0·36	0·06	0·08	0·15
0·50	0·14	0·43	0·08	0·16	0·18
0·90	0·23	0·48	0·12	0·29	0·22
0·99	0·36	0·50	0·18	0·45	0·25
Population average	0·15			0·17	0·19

TABLE X
Same case as Table IX.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0·04		35
0·05	0·07	−43	39
0·10	0·10	−3	36
0·20	0·17	15	31
0·30	0·24	20	27
0·40	0·31	22	24
0·50	0·39	21	21

TABLE XI
(Case 6)
$\alpha = 1$, $\beta = 1$, $\sigma = 1$, mean $n = 0.61$, $X/Z = 0.93$.
Full optimum for $X = Z - 0.013$: $x^0 = 0.25$, $F(x^0) = 0.35$.
Partial optimum (income-tax): $x_0 = 0.10$, $n_0 = 0.20$, $F(n_0) = 0.27$.

$F(n)$	x	y	$x(1-y)$	z	Full optimum x
0	0·10	0	0·10	0	0·25
0·10	0·10	0	0·10	0	0·25
0·50	0·14	0·15	0·11	0·06	0·28
0·90	0·32	0·41	0·19	0·54	0·44
0·99	0·90	0·49	0·46	1·84	0·62
Population average	0·18			0·20	0·32

TABLE XII
Same case as Table XI.

z	x	Average tax rate per cent	Marginal tax rate per cent
0	0·10		50
0·10	0·15	−50	58
0·25	0·20	20	60
0·50	0·30	40	59
1·00	0·52	48	57
1·50	0·73	51	54
2·00	0·97	51	52
3·00	1·47	51	49

Optimum Consumption Schedules
Cases 1-4

$\alpha=1, \beta=1, X/z=1.20$
$\alpha=1, \beta=0, X/z=1.10$
$\alpha=1, \beta=1, X/z=.99$
$\alpha=1, \beta=0, X/z=.93$

FIGURE 5

Optimum Consumption
Function

Distribution of
Consumption
after Tax

Distribution of Income
before Tax

Case 5:
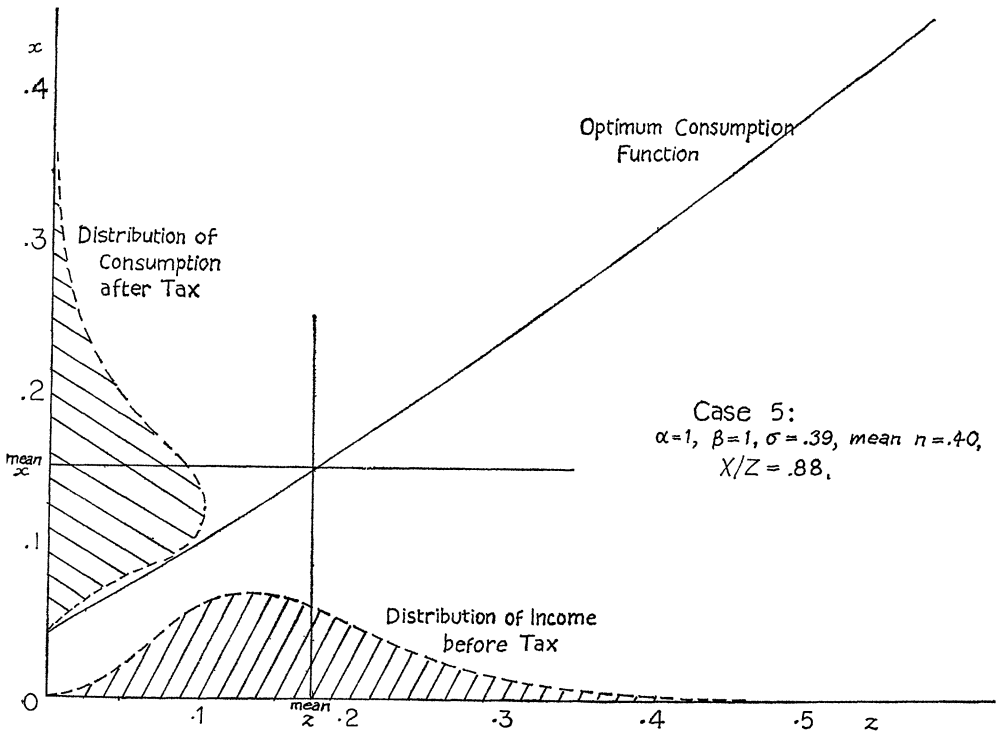$\alpha=1, \beta=1, \sigma=.39$, mean $n=.40$,
$X/Z=.88$.
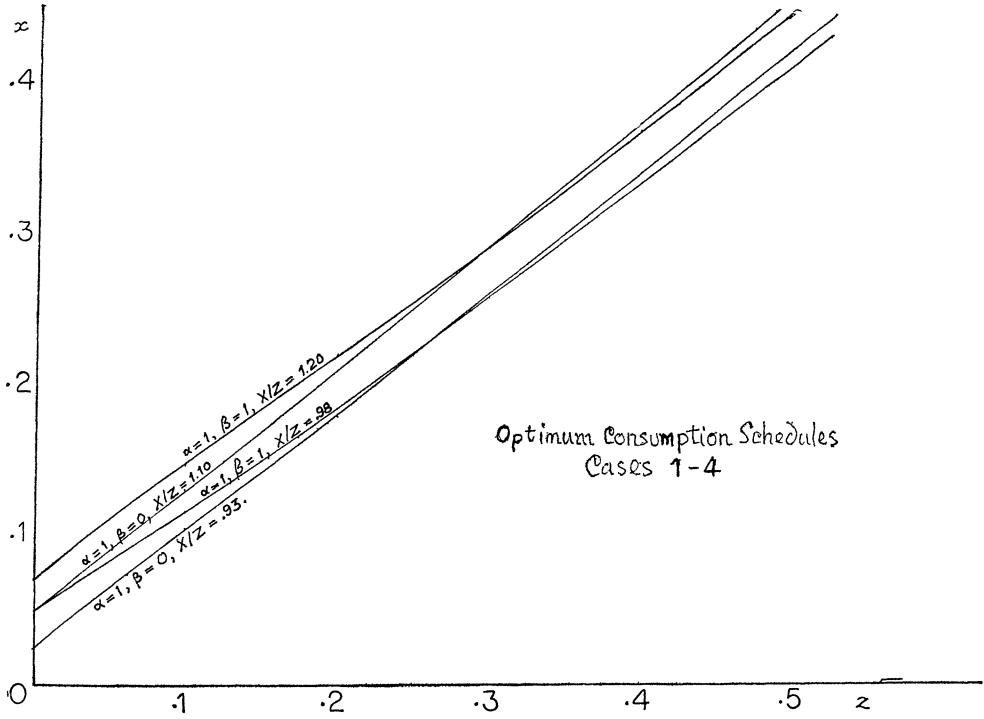
FIGURE 6

case, x_0, n_0, and the values of x, y and $x(1-y)$ (which measures utility) at the 10 per cent, 50 per cent, 90 per cent and 99 per cent points of the skill-distribution are given. In separate tables, the average and marginal tax rates are given for a representative range of values of z. Graphs of the optimal consumption schedule ($x = c(z)$) are given in Figs. 1 and 2. In Fig. 2, the distributions of x_n and z_n are displayed in case 5.

It will be noticed at once that, under the optimum regime, practically the whole population chooses to work in each of these cases: this contrasts, in some cases, with the full optimum, where sometimes a substantial proportion of the population is allowed to be idle. In most cases, a significant number work for less than a third of the time. It is also somewhat surprising that tax rates are so low. This means, in effect, that the income tax
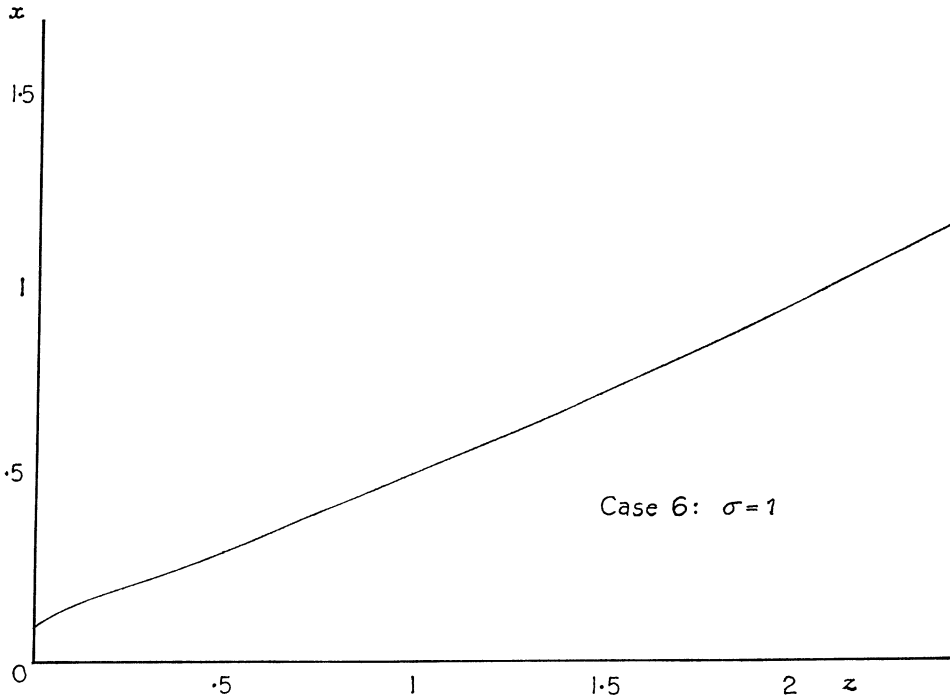

FIGURE 7

is not as effective a weapon for redistributing income, under the assumptions we have made, as one might have expected. It is not surprising that tax rates are higher when $\beta = 1$. When objectives are more egalitarian, more output is sacrificed for the sake of the poorer groups. Nevertheless, the difference between the optimum when only an income tax is available, and the full optimum, is rather large.

The examples have been chosen for X/Z fairly large: this corresponds to economies in which the requirements of government expenditure are largely met from the profits of public production, or taxation of private profits and commodity transactions. Tax rates are, as one might expect, fairly sensitive to changes in X/Z (i.e. to the production possibilities in the economy, and the extent to which income taxation is used to finance government expenditure as well as for " redistribution "). Tax rates are mildly sensitive to the choice of β. (When $\alpha = \frac{1}{2}$, the main features are unchanged).

Perhaps the most striking feature of the results is the closeness to linearity of the tax schedules. Since a linear tax schedule, which may be regarded as a proportional income tax in association with a poll subsidy, is particularly easy to administer, it cannot be said that the neglect of administrative costs in the analysis is of any importance, except that

considerations of administration might well lead an optimizing government to choose a perfectly linear tax schedule. The optimum tax schedule is certainly not exactly linear, however, and we have not explored the welfare loss that would arise from restriction to linear schedules: nevertheless, one may conjecture that the loss would be quite small. It is interesting, though, that in the cases for which we have calculated optimum schedules, the maximum marginal tax rate occurs at a rather low income level, and falls steadily thereafter.

This conclusion would not necessarily hold if the distribution of skills in the population had a substantially greater variance. The sixth case presented has $\sigma = 1$. So great a dispersion of known labouring ability does not seem to be at all realistic at present, but it is just conceivable if a great deal more were known to employers about the abilities of individual members of the population. The optimum is in almost all respects very different. Tax rates are high: a large proportion of the population is allowed to abstain from productive labour. The results seem to say that, in an economy where there is more intrinsic inequality in economic skill, the income tax is a more important weapon of public control than it is in an economy where the dispersion of innate skills is less. The reason is, presumably, that the labour-discouraging effects of the tax are more important, relative to the redistributive benefits, in the latter case.

10. CONCLUSIONS

The examples discussed confirm, as one would expect, that the shape of the optimum earned-income tax schedule is rather sensitive to the distribution of skills within the population, and to the income-leisure preferences postulated. Neither is easy to estimate for real economies. The simple consumption-leisure utility function is a heroic abstraction from a much more complicated situation, so that it is quite hard to guess what a satisfactory method of estimating it would be. Many objections to using observed income distributions as a means of estimating the distribution of skills will spring to mind. Yet the assumptions used in the numerical illustrations seem to fit observation fairly well, and are not in themselves implausible. It is not probable that work decisions are entirely, or even, in the long run, mainly, determined by social convention, psychological need, or the imperatives of cooperative behaviour: an analysis of the kind presented is therefore likely to be relevant to the construction and reform of actual income taxes.

Being aware that many of the arguments used to argue in favour of low marginal tax rates for the rich are, at best, premissed on the odd assumption that any means of raising the national income is good, even if it diverts part of that income from poor to rich, I must confess that I had expected the rigorous analysis of income-taxation in the utilitarian manner to provide an argument for high tax rates. It has not done so. I had also expected to be able to show that there was no great need to strive for low marginal tax rates on low incomes when constructing negative-income-tax proposals. This feeling has been to some extent confirmed. But my expectation that the minimum consumption level would be rather high has not been confirmed. Instead, virtually everyone is brought into the workforce. Since this conclusion is based on the analysis of an economy in which a man who chooses to work can work, I should not wish to see it applied in real economies. So long as there are periods when employment offered is less than the labour force available, one would perhaps wish to see the minimum income-level, assured to those who are not working, set at such a level that the number who choose not to work is as great as the excess of the labour force over the employment available. A rigorous analysis of this situation has still to be attempted. The results above do at least suggest that we should allow the least skilled to work for a substantially shorter period than the highly skilled.

I would also hesitate to apply the conclusions regarding individuals of high skill: for many of them, their work is, up to a point, quite attractive, and the supply of their labour

may be rather inelastic (apart from the possibilities of migration). There is scope for further theoretical work on this problem too. I conclude, for the present, that:—

(1) An approximately linear income-tax schedule, with all the administrative advantages it would bring, is desirable (unless the supply of highly skilled labour is much more inelastic than our utility function assumed); and in particular (optimal!) negative income-tax proposals are strongly supported.[1]

(2) The income-tax is a much less effective tool for reducing inequalities than has often been thought; and therefore

(3) It would be good to devise taxes complementary to the income-tax, designed to avoid the difficulties that tax is faced with. In the model we have been studying, this could be achieved by introducing a tax schedule that depends upon time worked (y) as well as upon labour-income (z): with such a schedule, one can obtain the full optimum, since one can, in effect, construct a different z-schedule for each n.[2] Such a tax would not be fully practicable, but we have other means of estimating a man's skill-level—such as the notorious I.Q. test: high values of skill-indexes may be sought after so much for prestige that they would not often be misrepresented. With any such method of taxation, the risks of evasion are, of course, quite great: but if it is true, as our results suggest, that the income tax is not a very satisfactory alternative, this objection must be weighed against the great desirability of finding some effective method of offsetting the unmerited favours that some of us receive from our genes and family advantages.

REFERENCES

[1]   Blum, W. J. and Kalven, H. Jr. *The Uneasy Case for Progressive Taxation* (University of Chicago Press, 1953).

[2]   Diamond, P. A. " Negative Income Taxes and the Poverty Problem—a Review Article ", *National Tax Journal* (September 1968).

[3]   Lydall, H. F. *The Structure of Earnings* (Oxford, 1968).

[4]   Mirrlees, J. A. " Characterization of the Optimum Income Tax " (unpublished).

[5]   Musgrave, R. A. *The Theory of Public Finance* (McGraw-Hill, 1959).

[6]   Shoup, C. S. *Public Finance* (Weidenfeld and Nicolson, 1969).

[7]   Vickrey, W. *Agenda for Progressive Taxation* (Ronald Press, N.Y., 1947).

    [1] The essential point of these proposals is that the marginal tax rate (as represented by rules for deductions from social security benefits) should be significantly less than 100 per cent. Proposals of this kind have sometimes been put forward in terms that suggest—quite wrongly of course—that any plausible-sounding negative income-tax proposal is better than a system in which all earnings are deducted from social security benefits. It was a major intention of the present study to provide methods for estimating desirable tax rates at the lowest income levels, and a surprise that these tax rates are the most difficult to determine, in a sense. They cannot be determined without at the same time determining the whole optimum income-tax schedule. To put things another way, no such proposal can be valid out of the context of the rest of the income-tax schedule.
    [2] I am indebted to Frank Hahn for pointing this out. It would seem to be true that lump-sum taxation is possible in any formal model where uncertainty is not introduced explicitly.